

STIC-Biotech/ChemLib

From: Hunt, Jennifer
Sent: Wednesday, April 10, 2002 10:12 AM
To: STIC-Biotech/ChemLib
Subject: Seq Search for 09/234,208

Please search and interference search SEQ ID NO:1 and 2 of 09/234,208.

Thanks,

Jennifer Hunt
Patent Examiner, Art Unit 1642
CM1-8D06 (mailbox 8E12)
(703)308-7548

Point of Contact:
Toby Port
Technical Info. Specialist
CM1 6A04
703-308-3534

RECEIVED
APR 10 2002
(STIC)

Searcher: _____
Phone: _____
Location: _____
Date Picked Up: 4/11
Date Completed: 4/12
Searcher Prep/Review: 10
Clerical: _____
Online time: 10

TYPE OF SEARCH:
NA Sequences: _____
AA Sequences: 2
Structures: _____
Bibliographic: _____
Litigation: _____
Full text: _____
Patent Family: _____
Other: _____

VENDOR/COST(where applic.)
STN: _____
DIALOG: _____
Questel/Orbit: _____
DRLink: _____
Lexis/Nexis: _____
Sequence Sys.: es
WWW/Internet: _____
Other (specify): _____

1. 100-1000000
2. 100-1000000
3. 100-1000000
4. 100-1000000
5. 100-1000000

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:28:18 ; Search time 40.19 Seconds
(without alignments)
145.603 Million cell updates/sec

Title: US-09-234-208b-1
Perfect score: 418
Sequence: 1 GTHSLRPAPVAVPVLRLMQP.....VGRGPDPAHVAVNLSTREG 79

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A.Geneseq_1101:*

- 1: /SID8/gcgdata/geneseq/AA1980.DAT:*
- 2: /SID8/gcgdata/geneseq/AA1981.DAT:*
- 3: /SID8/gcgdata/geneseq/AA1982.DAT:*
- 4: /SID8/gcgdata/geneseq/AA1983.DAT:*
- 5: /SID8/gcgdata/geneseq/AA1984.DAT:*
- 6: /SID8/gcgdata/geneseq/AA1985.DAT:*
- 7: /SID8/gcgdata/geneseq/AA1986.DAT:*
- 8: /SID8/gcgdata/geneseq/AA1987.DAT:*
- 9: /SID8/gcgdata/geneseq/AA1988.DAT:*
- 10: /SID8/gcgdata/geneseq/AA1989.DAT:*
- 11: /SID8/gcgdata/geneseq/AA1990.DAT:*
- 12: /SID8/gcgdata/geneseq/AA1991.DAT:*
- 13: /SID8/gcgdata/geneseq/AA1992.DAT:*
- 14: /SID8/gcgdata/geneseq/AA1993.DAT:*
- 15: /SID8/gcgdata/geneseq/AA1994.DAT:*
- 16: /SID8/gcgdata/geneseq/AA1995.DAT:*
- 17: /SID8/gcgdata/geneseq/AA1996.DAT:*
- 18: /SID8/gcgdata/geneseq/AA1997.DAT:*
- 19: /SID8/gcgdata/geneseq/AA1998.DAT:*
- 20: /SID8/gcgdata/geneseq/AA2000.DAT:*
- 21: /SID8/gcgdata/geneseq/AA2001.DAT:*
- 22: /SID8/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406	97.1	79	21	AAV97241
2	355	84.9	79	21	AAV97239
3	351	84.0	420	21	AAV97240
4	86	20.6	459	20	AAV31404
5	86	20.6	459	20	AAV31405
6	83.5	20.0	888	18	AAV19857
7	83.5	20.0	888	21	AAV18916
8	83.5	20.0	888	22	AAV12443
9	82	19.6	132	20	AAV31402
10	82	19.6	132	20	AAV31403
11	81.5	19.5	1257	15	AAV46627

12	78	18.7	122	20	AAV31386	HEV-US2 ORF3 prote
13	78	18.7	763	18	AAV31852	Mycoacterium tube
14	77	18.4	122	77	AAW93407	Swine HEV ORF 3 pr
15	77	18.4	329	22	AAV93933	Human protein sequ
16	75.5	18.1	198	21	AAV44537	Hepatitis C virus
17	75	17.9	530	21	AAV43300	Human ORFX ORF3064
18	75	17.9	1081	22	AAV95514	Human protein sequ
19	74.5	17.8	1246	20	AAV06296	Human transcriptio
20	74	17.7	616	21	AAV56941	Human prostate can
21	73.5	17.6	887	18	AAV19856	Rat semaphorin Z.
22	73	17.5	90	21	AAV54723	Arabidopsis thalia
23	73	17.5	148	21	AAV08181	Arabidopsis thalia
24	73	17.5	152	21	AAV54970	Arabidopsis thalia
25	73	17.5	164	21	AAV08180	Arabidopsis thalia
26	73	17.5	197	21	AAV08179	Arabidopsis thalia
27	73	17.5	639	22	AAV94220	Human protein sequ
28	73	17.5	1502	22	AAV39273	Human polypeptide
29	73	17.5	1565	22	AAV41059	Human polypeptide
30	72.5	17.3	833	21	AAV41835	Human ORFX ORF1599
31	72	17.2	122	20	AAV31383	HEV-US1 ORF3 prote
32	71.5	17.1	548	18	AAV07700	Human ETS2 repress
33	71.5	17.1	604	11	AAV06463	Derived protein of
34	71.5	17.1	604	20	AAV49848	Human pancreatic i
35	71.5	17.1	604	20	AAV33354	Human islet cell a
36	71.5	17.1	604	20	AAV80480	Islet cell antiod
37	71.5	17.1	604	21	AAV49322	Pancreatic islet c
38	71.5	17.1	615	22	AAV39081	Human polypeptide
39	71.5	17.1	2441	16	AAV79054	CREB binding prote
40	71.5	17.1	2441	19	AAV40058	Cellular transcrip
41	71.5	17.1	2441	21	AAV94252	Mouse nuclear CREB
42	71	17.0	148	21	AAV52235	Arabidopsis thalia
43	71	17.0	164	21	AAV52234	Arabidopsis thalia
44	71	17.0	197	21	AAV52233	Arabidopsis thalia
45	70.5	16.9	1026	22	AAV94315	Human protein sequ

ALIGNMENTS

RESULT 1
ID AAV97241
AAV97241 standard; Protein; 79 AA.
XX
AC AAV97241;
XX
DT 04-DEC-2000 (first entry)
XX
DE HER-2 C-terminal extracellular domain IIIA.
XX
KW HER-2: erbB-2; oncogene; receptor-like tyrosine kinase; insertion;
KW extracellular domain IIIA; antagonist; intron 8; C-terminal extension;
KW truncated HER-2; p66; dimerization inhibitor; cytosolic.
XX
OS Homo sapiens.
XX
PN WO200044403-A1.
XX
PD 03-AUG-2000.
XX
PF 20-JAN-2000; 2000WO-US01484.
XX
PR 20-JAN-1999; 99US-0234208.
XX
PA (UYOR-) UNITV OREGON HEALTH SCI.
XX
PI Doherty JK, Clinton GM, Adelman JP;
XX
DR WPI: 2000-499287/44.
XX
XX N-PSDB: AAA53783.
XX
PT Using polypeptides and antibodies that bind to the extracellular domain
PT of the receptor-like tyrosine kinase HER-2 to treat solid tumors of the
PT breast, lung, ovaries and colon

FT	/note= "N-myristoylation site"
FT	74...78
FT	/note= "N-glycosylation site"
FT	102...106
FT	/note= "cAMP- and GMP-dependent protein kinase phosphorylation site"
FT	155...159
FT	/note= "N-glycosylation site"
FT	167...171
FT	/note= "N-glycosylation site"
FT	189...195
FT	/note= "N-myristoylation site"
FT	291...295
FT	/note= "N-glycosylation site"
FT	318...339
FT	/note= "transmembrane domain"
FT	333...339
FT	/note= "N-myristoylation site"
FT	346
FT	/note= "Ala encoded by CCT"
FT	382...388
FT	/note= "N-myristoylation site"
FT	386...390
FT	/note= "N-glycosylation site"
FT	441...445
FT	/note= "N-glycosylation site"
FT	448...454
FT	/note= "N-myristoylation site"
FT	462...466
FT	/note= "N-glycosylation site"
FT	490...496
FT	/note= "N-myristoylation site"
FT	491...497
FT	/note= "N-myristoylation site"
FT	508...514
FT	/note= "N-myristoylation site"
FT	509...515
FT	/note= "N-myristoylation site"
FT	531...537
FT	/note= "N-myristoylation site"
FT	558...565
FT	/note= "N-myristoylation site"
FT	569...575
FT	/note= "N-myristoylation site"
FT	574...580
FT	/note= "N-myristoylation site"
FT	580...586
FT	/note= "N-myristoylation site"
FT	598...617
FT	/note= "transmembrane domain"
FT	610...616
FT	/note= "N-myristoylation site"
FT	643...649
FT	/note= "N-myristoylation site"
FT	663...669
FT	/note= "N-myristoylation site"
FT	666...672
FT	/note= "N-myristoylation site"
FT	667...673
FT	/note= "N-myristoylation site"
FT	668...674
FT	/note= "N-myristoylation site"
FT	669...675
FT	/note= "N-myristoylation site"
FT	670...676
FT	/note= "N-myristoylation site"
FT	686...674
FT	/note= "N-myristoylation site"
FT	879...885
FT	/note= "N-myristoylation site"
XX	W0200056889-A2.
XX	

```

PD 28-SEP-2000.
XX
PF 01-MAR-2000; 2000WO-US05601.
XX
PR 23-MAR-1999; 99US-0125774.
PR 23-MAR-1999; 99US-0125778.
PR 24-MAR-1999; 99US-0125826.
PR 31-MAR-1999; 99US-0127035.
PR 05-APR-1999; 99US-0127706.
PR 21-APR-1999; 99US-0130359.
PR 27-APR-1999; 99US-0131270.
PR 27-APR-1999; 99US-0131272.
PR 27-APR-1999; 99US-0131291.
PR 04-MAY-1999; 99US-0132371.
PR 04-MAY-1999; 99US-0132379.
PR 04-MAY-1999; 99US-0132383.
PR 25-MAY-1999; 99US-0135750.
PR 08-JUN-1999; 99US-0138166.
PR 20-JUL-1999; 99US-0144791.
PR 03-AUG-1999; 99US-0146970.
PR 09-DEC-1999; 99US-0170262.
XX
XX
PA (GENE ) GENENTECH INC.
PI Desnoyers L, Eaton DL, Goddard A, Godowski PJ, Gurney AL, Pan J;
PI Stewart TA, Watanabe CK, Wood WL, Zhang Z;
PI
DR N-PSDB; AAA65343.
XX
XX
XX WPI: 2000-628263/60.
DR
XX
XX
XX Novel secreted and transmembrane polypeptides useful for diagnosing
PT tumour in a mammal, for identifying agonists and antagonists of the
PI polypeptide and for therapeutic use
XX
XX
PS Claim 12; Fig 16; 222pp; English.
XX
XX
XX The present sequence represents a secreted or transmembrane polypeptide.
CC
CC The specification describes polypeptides designated PRO1484, PRO4334,
CC PRO1122, PRO1889, PRO1890, PRO1887, PRO1785, PRO4353, PRO4357, PRO4405,
CC PRO4356, PRO4352, PRO4380, PRO4354, PRO4408, PRO5737, PRO4425, PRO5990,
CC PRO6030, PRO4424, PRO4422, PRO4430 and PRO4499. PRO1889 polypeptide is
CC useful for diagnosing tumour in a mammal. The polypeptides, their
CC agonists and antagonists are useful treating a condition associated with
CC expression or activity of the polypeptide. Conditions treated include
CC obesity, diabetes or hyper- or hypo-insulinemia. The polypeptides are
CC capable of inducing proliferation of mammalian kidney mesangial cells
CC and are therefore useful for treating kidney disorders associated with
CC decreased mesangial cell function such as Berger's disease or other
CC nephropathies associated with Schönlein-Henoch purpura, celliac disease,
CC dermatitis herpetiformis or Crohn's disease. The nucleic acids may be used
CC to generate transgenic animals for use in development and screening of
CC therapeutically useful reagents and also for chromosome identification
CC and tissue typing.
XX
XX
XX Sequence 888 AA;
SQ

```

XX AAU12443;
AC XX
XX 24-OCT-2001 (first entry)
DE Human PRO4353 polypeptide sequence.
XX
XX Human secretory and transmembrane; PRO: mammalian; cancer; lung;
KW breast; prostate; cervical; tumour necrosis factor-alpha; TNF-alpha;
KW cartilage; ear; proliferation; glucose; free fatty acid; skeletal muscle;
KW adipocyte; A-peptide; factor VIIA; gene therapy.
XX
OS Homo sapiens.
XX
PN WO200140466-A2.
XX
PD 07-JUN-2001.
XX
PF 01-DEC-2000; 2000MO-US32678.
XX
PR 01-DEC-1999; 99MO-US28301.
PR 01-DEC-1999; 99MO-US28634.
PR 02-DEC-1999; 99MO-US28551.
PR 02-DEC-1999; 99MO-US28564.
PR 02-DEC-1999; 99MO-US28565.
PR 09-DEC-1999; 99US-0170262.
PR 16-DEC-1999; 99MO-US30095.
PR 20-DEC-1999; 99MO-US30911.
PR 20-DEC-1999; 99MO-US30939.
PR 30-DEC-1999; 99MO-US31243.
PR 06-JAN-2000; 2000MO-US00277.
PR 11-FEB-2000; 2000MO-US00376.
PR 18-FEB-2000; 2000MO-US04341.
PR 18-FEB-2000; 2000MO-US04342.
PR 22-FEB-2000; 2000MO-US04414.
PR 24-FEB-2000; 2000MO-US04914.
PR 24-FEB-2000; 2000MO-US05004.
PR 01-MAR-2000; 2000MO-US05601.
PR 20-MAR-2000; 2000MO-US07377.
PR 21-MAR-2000; 2000MO-US07532.
PR 30-MAR-2000; 2000MO-US08439.
PR 17-MAY-2000; 2000MO-US13705.
PR 22-MAY-2000; 2000MO-US14042.
PR 30-MAY-2000; 2000MO-US14941.
PR 02-JUN-2000; 2000MO-US15264.
PR 10-NOV-2000; 2000MO-US30873.
XX
XX (GETH) GENENTECH INC.
PA
PI Baker KP, Beresini M, DeGeorge L, Desnoyers L, Filvaroff E, Gao W;
PI Geritsen ME, Goddard A, Godowski PJ, Gurney AL, Sherwood S;
PI Smith V, Stewart TA, Tumas D, Watanabe CK, Wood WI, Zhang Z;
XX
XX WPI; 2001-408281/43.
DR N-PSDB: AAS21515.
XX
XX Isolated, secretory and transmembrane PRO polypeptide used to detect
PT other PRO polypeptides, link bioactive molecules to cells expressing
PT PRO polypeptides, and detect the presence of mammalian tumours e.g.
PT lung, breast, prostate, cervical
XX
XX Claim 12; Fig 544; 813pp: English.
PS
XX AAU12172-AAU12446 represent novel human secretory and transmembrane
CC PRO polypeptides. The PRO polypeptides are useful to detect other
CC PRO polypeptides, to link bioactive molecules to cells expressing
CC PRO polypeptides, to modulate biological activities of cells expressing
CC PRO polypeptides, and to detect the presence of mammalian lung, colon,
CC breast, prostate, rectal, cervical or liver tumours by comparing PRO
CC polypeptide expression in a cell sample to that in a control sample.
CC Some of the 275 sequences are also useful to stimulate the release of
CC tumour necrosis factor-alpha (TNF-alpha) from human blood, the

CC proliferation or differentiation of chondrocytes, the proliferation or
CC gene expression in pericyte cells, the release of proteoglycans from
CC cartilage, the proliferation of inner ear utricular supporting cells or
CC of T-lymphocytes, the release of a cytokine from peripheral blood
CC monocytes (PBMCs), or the proliferation of endothelial cells. Some of
CC the PRO polypeptides may modulate glucose or free fatty acid uptake by
CC skeletal muscle cells or by adipocytes; or inhibit binding of A-peptide
CC to factor VIIa. The PRO polypeptides can be used in assays to identify
CC molecules involved in binding interactions. The polynucleotides encoding
CC PRO polypeptides can be used to generate probes, antisense RNA/DNA,
CC transgenic or knock out animals and can be used in gene therapy.
XX
SQ Sequence 888 AA:
XX
Query Match 20.0%; Score 83.5; DB 22; Length 888;
Best Local Similarity 35.4%; Pred. No. 0.76;
Matches 28; Conservative 7; Mismatches 33; Indels 11; Gaps 5;
QY 1 GTH---SLPPRAAVPVLRMQGPAPHVLSFLRP-SMD---LVSAPYSLLAPLSPT 51
Db 698 gpndldsglllppeqtpipqkrlpvp-nphnhaigrwdhnp1lpasasslllllapa 756
QY 52 SVPISEVSVGRGPPDPDAHV 70
Db 757 rapqppapage-plpogrl 774
RESULT 9
AAV31402
ID AAV31402 standard; Protein; 132 AA.
XX
AC AAV31402;
XX
DT 12-OCT-1999 (first entry)
XX
DE HEV US and CMP-KDO synthetase (CKS) fusion protein.
XX
XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
KW vaccine; passive immunisation.
KW
OS Hepatitis E virus.
OS
XX WO9919732-A1.
PN
XX 22-APR-1999.
PD
XX 15-OCT-1998; 98MO-US21941.
PF
XX 15-OCT-1997; 97US-0061199.
PR
XX (ABBO) ABBOTT LAB.
PA
PI Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GG;
PI WPI; 1999-288017/24.
DR N-PSDB: AAZ00278.
XX
PT Detection of United States isolates of hepatitis E virus
PT
PS Example 10; Page 240; 260pp: English.
XX
XX The invention provides a method for detecting a US (sub)type hepatitis E
CC virus (US-HEV), or its naturally occurring variants in a sample by
CC treatment with a binding partner specific for a marker of the virus, and
CC then detecting any complex formed. The method is used to diagnose
CC infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
CC open reading frames (ORF) in US-HEV and host cells expressing these ORFs
CC are useful in vaccines or for passive immunisation. The polypeptides are
CC also used to raise specific antibodies (useful as immunoassay reagents).
CC Fragments of nucleic acid from US-HEV are useful as primers and probes in
CC usual hybridisation and amplification assays for detecting infection.
XX


```

FT Modified-site 857 /note= "O-linked glycosylation site"
FT Modified-site 859 /note= "O-linked glycosylation site"
FT Modified-site 861 /note= "O-linked glycosylation site"
FT Modified-site 863 /note= "O-linked glycosylation site"
FT Modified-site 905 /note= "O-linked glycosylation site"
FT Modified-site 913 /note= "O-linked glycosylation site"
FT Peptide 933..942 /note= "O-linked glycosylation site"
FT Binding-site 944..945 /note= "used for primer design"
FT Modified-site 950 /note= "Chondroitin sulphate attachment site"
FT Modified-site 967 /note= "O-linked glycosylation site"
FT Modified-site 1136..1147 /note= "N-linked glycosylation site"
FT Peptide /note= "used in primer design"
FT Modified-site 1164 /note= "N-linked glycosylation site"
PN WO9403601-A.
PD 17-FEB-1994.
XX
XX 03-AUG-1993; 93WO-US07306.
XX
XX 03-AUG-1992; 92US-0922911.
XX
XX (UYNV ) UNITV NEW YORK STATE.
XX
XX Margolis RK, Margolis RU, Rauch U;
XX
XX MPI: 1994-065690/08.
XX
XX N-PSDB; AAQ57710.
XX
XX Eukaryotic neurocan polypeptide(s) with epidermal growth factor,
XX lectin or complement binding activity - used in the diagnosis,
XX treatment or research of hypersensitivity and allergic diseases
XX
XX Claim 1; Page 69-75; 105pp; English.
XX
XX This sequence represents a neurocan polypeptide. This protein has
XX several biological activities, including cell adhesion, leukocyte-
XX endothelial cell recognition, tissue-related inflammation allergies,
XX cellular and/or humoral hypersensitivity, trauma, neuronal
XX development, and cell transport and/or infection. Compositions
XX containing them can be used as modulators of these conditions, and
XX may be used as therapeutic, diagnostic, and/or research tools.
XX Neurocan peptides can be used to mimic proteins, such as lectins,
XX cell adhesion molecules, versicans, aggrecans or gelsolins, as
XX receptor or effector subtypes. The protein can be used to treat
XX diseases involving a qualitative or quantitative pathological
XX abnormality of cell adhesion or leukocyte-endothelial cell recognition,
XX or a functionally associated molecule such as a membrane cytoplasmic
XX protein, lipid, carbohydrate, saccharide, nucleoside, enzyme or ion.
XX
XX Sequence 1257 AA:
SQ

```

Query Match 19.5%; Score 81.5; DB 15; Length 1257;
 Best Local Similarity 35.0%; Pred. No. 1.9;
 Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

```

QY 4 SLLPRRAAVPVLPMQPG---FAHPVLSFLR-----PSMDIVSAFYSLPLAPLS--PT 51
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 610 ssipsealseavsqspgdpdipvmrlapklwlpshstlvnpvslpispasppls 669
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 52 SVP-----ISPVSGRGPPDP 67

```

```

DB 670 svpeegavrvpsfg-aedpe 688
   ||| : ||| | ||:
RESULT 12
AAV31386
ID AAV31386 standard; Protein; 122 AA.
XX
XX AAV31386;
AC
XX
XX 12-OCT-1999 (first entry)
DE
XX HEV-US2 ORF3 protein.
XX
XX Hepatitis E virus; HEV; binding partner; virus; US-HEV infection;
XX vaccine; passive immunisation.
XX
XX Hepatitis E virus.
OS
XX
XX Key Location/Qualifiers
FH Misc-difference 97 /label= unknown
FT /note= "encoded by GYT"
FT
FT
PN WO9919732-A1.
PD 22-APR-1999.
XX
XX 15-OCT-1998; 98WO-US21941.
XX
XX 15-OCT-1997; 97US-0061199.
XX
XX (ABBO ) ABBOTT LAB.
XX
XX Dawson GJ, Desai SM, Erker JC, Mushahwar IK, Schlauder GC;
XX
XX MPI: 1999-288017/24.
XX
XX N-PSDB; AAO0267.
XX
XX Detection of United States isolates of hepatitis E virus
XX
XX Claim 18; Page 223; 260pp; English.
XX
XX The invention provides a method for detecting a us (sub)type hepatitis E
XX virus (US-HEV), or its naturally occurring variants in a sample by
XX treatment with a binding partner specific for a marker of the virus, and
XX then detecting any complex formed. The method is used to diagnose
XX infection with US-HEV. Polypeptides from US-HEV, antibodies specific for
XX open reading frames (ORF) in US-HEV and host cells expressing these ORFs
XX are useful in vaccines or for passive immunisation. The polypeptides are
XX also used to raise specific antibodies (useful as immunoassay reagents).
XX Fragments of nucleic acid from US-HEV are useful as primers and probes in
XX usual hybridisation and amplification assays for detecting infection. The
XX present sequence represents a HEV-US2 ORF3 protein.
XX
XX Sequence 122 AA:
SQ

```

Query Match 18.7%; Score 78; DB 20; Length 122;
 Best Local Similarity 36.1%; Pred. No. 0.31;
 Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 3;

```

QY 1 GTRSLPRRAAVPVLPMQPGFAHPVLSFLRPSMDIVSAFYSLPLAPL---SPTSVPISP 57
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 55 gvtglllspss--pspdlqtpss-pmstfhngylelaldsdaplxpvgvtspspplpp 111
   | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 58 V 58
DB 112 V 112

```

RESULT 13
 AAM31852

ID		AAM31852 standard; Protein; 763 AA.	
XX			
AC		AAM31852;	
XX			
DT		27-APR-1998 (first entry)	
XX			
DE	Mycobacterium tuberculosis 74 kda protein.		
XX			
KW	Tuberculosis; mycobacteria; infection; diagnosis; antimicrobial; antibiotic; vaccine.		
XX			
OS	Mycobacterium tuberculosis.		
XX			
PX	MO9741252-A2.		
PD			
PP	06-NOV-1997.		
PR	18-APR-1997; 97WO-EPO1973. 29-APR-1996; 96DE-4017184.		
PA	(GBFB) GBF GES BIOTECHE FORSCHUNG GMBH.		
PI	Espitia C, Honisch C, Moreno C, Singh M;		
DR	WPJ: 1997-549750/50.		
N-PSDB:	AAT93610.		
PT	New DNA and related proteins or RNA derived from M. tuberculosis - used for diagnosis of mycobacterial infections, monitoring vaccination and development of anti-mycobacterial agents		
PS	Claim 5; Fig 13; 55pp; English.		
CC	This novel 74 kDa protein is encoded by an open reading frame of a Mycobacterium tuberculosis DNA fragment (see AAT93610) containing polymorphic GC-rich sequences. Its amino acid sequence shows a high proline content, but there is no homology to any known proline-rich antigens of mycobacteria. Novel M. tuberculosis proteins (see AAM31851-57) are claimed. These can be produced as recombinant proteins, especially in bacterial, yeast, fungal or higher eukaryote host cells, and used for diagnosing tuberculosis and other mycobacterial infections in humans or animals. The claimed proteins can also be used for epidemiological studies, for monitoring vaccination, and for the development of vaccines and anti-mycobacterial drugs.		
SQ	Sequence 763 AA;		
Query Match	18.7%; Score 78; DB 18; Length 763;		
Best Local Similarity	34.8%; Pred No. 2.5; Indels 8; Gaps 2		
Matches	24; Conservative 4; Mismatches 33;		
OY	4 SLLEPAAVPVPLRMQPAPHVLISFLRPMDLVSFAFYSLPLAIPSP---TSVIPSPYS 59 : : : Db 9 alaplpapppapaepekxpfpappapccmllvsaaipcpappapapkaptprvp 68		
OY	60 VGRGDPDA 68 Db 69 ----pappa 73		
RESULT_14			
ID	AAM93407 standard; Protein; 122 AA.		
XX			
AC	AAM93407;		
XX			
DT	11-JUN-1999 (first entry)		
DE	Swine HEV ORF 3 protein.		

[illegible]

```

XX 28-JUL-2000; 2000EP-0116126.
PF
XX 29-JUL-1999; 99JP-0248036.
PR 27-AUG-1999; 99JP-0300253.
PR 11-JAN-2000; 2000JP-0118776.
PR 02-MAY-2000; 2000JP-0183767.
PR 09-JUN-2000; 2000JP-0241899.
XX
XX (HELI-) HELIX RES INST.
XX Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
XX
XX WPI; 2001-318749/34.
DR
XX
XX Primer sets for synthesizing polynucleotides, particularly the 5602
PT full-length cDNAs defined in the specification, and for the detection
PT and/or diagnosis of the abnormality of the proteins encoded by the
PT full-length cDNAs -
XX
XX PS Claim 8; SEQ ID 13931; 2537pp + CD ROM; English.
XX
XX CC The present invention describes primer sets for synthesizing 5602
CC full-length cDNAs defined in the specification. Where a primer set
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
CC to the complementary strand of a polynucleotide which comprises one of
CC the 5602 nucleotide sequences defined in the specification, where the
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
CC of an oligonucleotide comprising a sequence complementary to the
CC complementary strand of a polynucleotide which comprises a 5'-end
CC sequence and an oligonucleotide comprising a sequence complementary to a
CC polynucleotide which comprises a 3'-end sequence, where the
CC oligonucleotide comprises at least 15 nucleotides and the combination of
CC the 5'-end sequence/3'-end sequence is selected from those defined in
CC the specification. The primer sets can be used in antisense therapy and
CC in gene therapy. The primers are useful for synthesizing polynucleotides,
CC particularly full-length cDNAs. The primers are also useful for the
CC detection and/or diagnosis of the abnormality of the proteins encoded by
CC the full-length cDNAs. The primers allow obtaining of the full-length
CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
CC represent oligonucleotides, all of which are used in the exemplification
CC of the present invention.
XX
XX SO Sequence 329 AA:

```

Query Match 18.4%; Score 77; DB 22; Length 329;
 Best Local Similarity 34.3%; Pred. No. 1.2;
 Matches 24; Conservative 8; Mismatches 24; Indels 14; Gaps 2;

```

QY 8 RPAAVPYLRMPGPAHVPVLSFLRPSWDLVSAFYSLPLAPLSPTSVPSVSGRGDPD 67
   ||: | : |||| | ||| | ||| | : | ||: |
Db 231 rpskydvrgiqkpgpak-----vppkslappla--svpsapsapgppepp 276
QY 68 AHYAVNLSRY 77
   | : : | |
Db 277 aslstntpey 286

```

Search completed: April 11, 2002, 09:29:08
 Job time: 50 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:28:23 ; Search time 25.01 Seconds
(without alignments)
240.615 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418
Sequence: 1 GFRSLPRAAVPPLRMQP.....VGRGPDPAHVAVNLSRYEG 79

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: PIR.*
2: PIR1.*
3: PIR2.*
4: PIR3.*
5: PIR4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	81.5	19.5	1257	2 S28764	neurocan precursor
2	81	19.4	200	2 H84715	probable phytoecyan
3	81	19.4	1006	2 G66292	hypothetical prote
4	80	19.1	1834	1 JDMU1	DNA-directed RNA p
5	80	19.1	1840	2 G85422	hypothetical prote
6	78	18.7	249	2 S72619	hypothetical prote
7	77.5	18.5	217	2 T51031	related to finger
8	77	18.4	503	2 T19319	hypothetical prote
9	76.5	18.3	1110	2 T19673	hypothetical prote
10	76	18.2	356	2 A96826	hypothetical prote
11	76	18.2	848	2 S48273	genome-transcrip
12	75	17.9	189	2 S49600	probable transcrip
13	75	17.9	1952	2 T48814	beta-adaptin homol
14	74.5	17.8	894	2 T13029	hypothetical prote
15	74	17.5	1520	2 T10273	hypothetical prote
16	73	17.3	518	2 F70831	probable PPE prote
17	72	17.2	440	2 A44081	kappa-type oploid
18	72	17.2	1914	2 T42635	tenascin Y precurs
19	71.5	17.1	2441	2 S39161	CREB-binding prote
20	71	17.0	906	2 A71438	probable resistanc
21	70.5	16.9	291	2 S27721	hypothetical prote
22	70.5	16.9	443	2 T27877	hypothetical prote
23	70.5	16.9	788	1 JDLVHH	DNA-directed DNA p
24	70.5	16.9	1268	1 S52781	neurocan - mouse
25	70	16.7	491	2 S41182	DNA-directed RNA p
26	70	16.7	650	2 S41181	DNA-directed RNA p
27	70	16.7	954	2 E86174	protein F1919.26
28	70	16.7	2187	2 T30826	nascent polypeptid
29	70	16.7	2318	2 S45306	notch 3 protein -

30	70	16.7	2321	2 S78549	notch3 protein - h
31	69.5	16.6	433	2 T09284	TEA domain-contain
32	69.5	16.6	608	2 A46312	gag polyprotein -
33	69.5	16.6	684	2 T25603	hypothetical prote
34	69	16.5	148	2 T38881	caudal-type homeot
35	69	16.5	280	2 A75526	S-layer-like array
36	69	16.5	1121	2 T02764	myosin-I binding p
37	69	16.5	1353	2 T00249	hybrid proline-ric
38	68.5	16.4	401	2 T51407	arabinogalactan-pr
39	68.5	16.4	461	2 T10285	probable spermidin
40	68.5	16.4	554	2 T36345	probable potassium
41	68.5	16.4	1017	2 T31354	CDA peptide synthet
42	68.5	16.4	7463	2 T36248	hybrid proline-ric
43	68	16.3	301	2 J01663	hypothetical prote
44	68	16.3	444	2 E83802	myosin VII7 - human
45	68	16.3	543	2 S35047	

ALIGNMENTS

```
RESULT 1
S28764
neurocan precursor - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 22-Nov-1993 #sequence_revision 01-Sep-1995 #text_change 04-Feb-2000
C:Accession: S28764
R:Rauch, U.; Karthikeyan, L.; Maurel, P.; Margolis, R.U.; Margolis, R.K.
J. Biol. Chem. 267, 19536-19547, 1992
A:Title: Cloning and primary structure of neurocan, a developmentally regulated, aggr
A:Reference number: S28764; MUID:92406907
A:Accession: S28764
A:Molecule type: mRNA
A:Residues: 11257 <RNU>
A:Cross-references: EMBL:M97161; NID:9205649; PIDN:AC37679.1; PID:9205650
C:Superfamily: aggrecan; C-type lectin homology; complement factor H repeat homology;
C:Keywords: chondroitin sulfate proteoglycan; glycoprotein
F:1-22/Domain: signal sequence #status predicted <MAT>
F:23-1257/Product: neurocan #status predicted <MAT>
F:176-253/Domain: link protein repeat homology <LNK1>
F:274-355/Domain: link protein repeat homology <LNK2>
F:364-366/Region: cell attachment (R-G-D) motif
F:953-984/Domain: EGF homology <EGF>
F:1029-1149/Domain: C-type lectin homology <LCH>
F:1156-1212/Domain: complement factor H repeat homology <RHD>
F:1121,339,737,967,1164/Binding site: carboxylate (Asn) (covalent) #status predicted
F:372,410/Binding site: chondroitin sulfate (Ser) (covalent) #status predicted
F:944/Binding site: chondroitin sulfate (Ser) (covalent) #status experimental

Query Match 19.5% Score 81.5; DB 2; Length 1257;
Best local similarity 35.0%; Pred. No. 2.7;
Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

OY 4 SLPPRAAVPPLRMQP---PAHPVLSFLR-----PSMDVSAFSLPLAPLS--PT 51
DB 610 SSIPSEALSAVSLQSPDGSPPRIYAMLRKMLPHSLIVNVSPIPLSPAPLS 669
OY 52 SVP-----ISPVSVGRGPPD 67
DB 670 SVPEQAVRPVSFG-AEDPE 688

RESULT 2
H84715
probable phytoecyan [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 02-Feb-2001
C:Accession: H84715
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Unayam, L.; Tallon,
euss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter
Nature 402, 761-768, 1999
```

```

A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: AB4420; MUID:20083487
A:Accession: H84715
A>Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-200 <STD>
A:Cross-references: GB:AEO05172; NID:g3746072; PIDN:AAC63847.1; GSPDB:GN00139
C:Genetics:
A:Gene: At2g31050
A:Map position: 2

Query Match          19.4%; Score 81; DB 2; Length 200;
Best Local Similarity 34.3%; Pred. No. 0.37;
Matches    23; Conservative   5; Mismatches    23; Indels    16; Gaps     2;

OY      9 PAAVPPRLMOPGPAHPVLFLRPSMDLVSATFSLPLAPLSPTSVIPISVSGRGPDPA 68
         | | | | | | | | | | | | | | | | | | : | : | | | | | |
Db       132 PVAAIV-----PGVPRPSSFSFPS-----GSPLAESPVNHAFVQXQGKPSAP 175

OY      69 HVAVNLNS 75
         | | | |
Db       176 HSASNS 182

RESULT      3
G86292
hypothetical protein AAF82153.1 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: G86292
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federpiel, N.A.; Kaul, S.; White, O.; Alonso,
Chlin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.N.;
ansen, N.F.; Hughes, B.; Huizart, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: AB6141; MUID:21016719
A:Accession: G86292
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1006 <STD>
A:Cross-references: GB:AEO05172; NID:g8927662; PIDN:AAF82153.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1

Query Match          19.4%; Score 81; DB 2; Length 1006;
Best Local Similarity 31.4%; Pred. No. 2.3;
Matches    22; Conservative   5; Mismatches    21; Indels    22; Gaps     2;

OY      6 LRRPAAVPPLRMOGCPAHNVLSFLRPSMDLVSAFSLPLAPLSPTSVIPISVSGRGPD 65
         ||| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db       336 LPPEATLPPLPLPPPP-----SLPVPICSP---PPPVIINGCAP 373

OY      66 PDAAHVANLNS 75
         | | | |
Db       374 PPCVTCTVOYS 383

RESULT      4
JDMU1
DNA-directed RNA polymerase (EC 2.7.7.6) II largest chain - Arabidopsis thaliana
N:Alternate names: DNA-directed RNA polymerase II 20Sk chain; protein F4B14.70
C:Species: Arabidopsis thaliana (mouse-ear cross)
C>Date: 31-Mar-1993 #sequence_revision 11-Jun-1999 #text_change 20-Aug-1999
C:Accession: T04690; S12071; S27346; S11960
S:Beran, M.; Rose, M.; Hempel, S.; Entian, K.D.; Heiseisel, J.; Mewes, H.W.; Meyer, K.F.)
submitted to the Protein Sequence Database, October 1998

```

A:Reference number: 215380
 A:Accession: R04690
 A:Molecule type: DNA
 A:Residues: 1-1834 <BEV>
 A:Cross-references: EMBL:AL031986
 A:Experimental source: cultivar Columbia; BAC clone FAB14
 R:Navrath, C.; Schell, J.; Koncz, C.
 Mol. Gen. Genet. 223, 65-75, 1990
 A:Title: Homologous domains of the largest subunit of eucaryotic RNA polymerase II an
 A:Reference number: SI2071; MUID:91080867
 A:Accession: SI2071
 A:Molecule type: DNA
 A:Residues: 1-421,'S','423-732','D','734-1055','R','1057-1714','SPSPSY','1715-1834 <NAMI>
 A:Cross-references: EMBL:X52954; NID:g16504; PIDN:CAA37130.1; PID:g16505
 A:Experimental source: cv. Columbia
 A>Note: the authors translated the codon AGC for residue 1755 as Arg
 A:Accession: S27346
 A:Molecule type: mRNA
 A:Residues: 510-732,'D','734-1055','R','1057-1714','SPSPSY','1715-1834 <NAMI>
 A:Experimental source: cv. Columbia
 R:Dietch, M.A.; Prenger, J.P.; Gullioyle, T.J.
 Plant Mol. Biol. 15, 207-223, 1990
 A:Title: Analysis of the genes encoding the largest subunit of RNA polymerase II in A
 A:Reference number: SI1960; MUID:91355869
 A:Accession: SI1960
 A:Molecule type: DNA
 A:Residues: 1-116,125-192,'NSKEE','198-297','R','299-302','R','304-400','KE','403','VDYGP
 A:Cross-references: EMBL:X52494; NID:g16493; PIDN:CAA36735.1; PID:g16494
 A>Note: the authors translated the codon CCT for residue 1083 as Ala
 C:Genetics:
 A:Gene: rpl1215; RPB1
 A:Map position: 4
 A:Introns: 28/3; 123/3; 218/3; 271/3; 325/3; 411/1; 440/3; 552/3; 648/2; 734/3; 1755/
 A>Note: FAB14.70
 C:Superfamily: human DNA-directed RNA polymerase II largest chain
 C:Keywords: DNA binding; nucleotidyltransferase; phosphoprotein; tandem repeat; trans
 F:66-109/Region: zinc finger CCCC motif
 F:1531-1812/Region: 7-residue repeats

```

Query March 19.1%; Score 80; DB 2; Length 1840;
Best Local Similarity 36.4%; Pred. No. 5.9;
Matches 28; Conservative 3; Mismatches 36; Indels 10; Gaps 3;

Oy 7 PRPAAPVPLRMQP---GRAPVLSFLRPSMDLVSAPFSLPLAPLSPTSPVPSFVSGR 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1736 PSIAVSPSNARLSPASPSPTSPNVSPTSPSPSPSPSPS---PSPPTSPSPSSGA 1791
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Oy 63 GPD--PDAHVAVNLSRY 77
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 1792 SPDYSPSAGYSPTLPGY 1808
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 6
S72619
hypothetical protein 1a - anthracnose fungus (Colletotrichum gloeosporioides) retrotrans
C:Species: Glomerella cingulata, Colletotrichum gloeosporioides
C:Date: 24-Oct-1998 #sequence_revision 24-Oct-1998 #text_change 17-Mar-1999
C:Accession: S72619
R:He, C.; Nourse, J.P.; Kelemu, S.; Irwin, J.A.G.; Manners, J.M.
Mol. Gen. Genet. 252, 320-331, 1996
A:Title: CgT1: a non-ITR retrotransposon with restricted distribution in the fungal phyto
A:Reference number: S72619; MUID:96439839
A:Accession: S72619
A:Molecule type: DNA
A:Residues: 1-249 <HEA>
A:Cross-references: EMBL:L76172
A:Experimental source: biotype B, isolate 0062
A:Note: in the authors' translation residues 1-10 are not shown
C:Genetics:
A:Module element: retrotransposon CgT1

```

[illegible]

```

Query Match          18.7% ; Score 78 ; DB 2 ; Length 249 ;
Best Local Similarity 34.7% ; Pred. No. 0.94 ;
Matches 25 ; Conservative 7 ; Mismatches 26 ; Indels 14 ; Gaps 3 ;

QY      5 LLPR-----PAAVPLRLMQGCPAHFVLFLRPSW---DLVSATYSILPLAPLSPTSVP 54
           |||         | | : | | | | | | | | | | | | | | | | | | | | | |
Db      24 LVPRNVCHPRNTNTAQPTQRQREPSAMASQPAAGSOGPIOLLSSMHLKPTTPPPPTSLP 83

OY      55 ISPVSVCGRPDP 66
           - - - - - | | | | |
Db      84 PRP----RGLTTP 91

RESULT      7
T51031
related to finger protein XFG 68 [imported] - Neurospora crassa
N:Alternate names: protein B15120..10
C:Species: Neurospora crassa
C>Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 21-Jul-2000
C:Accession: T51031
R:Schultze, U.; Alyn, V.; Hohenisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nakamura,
submitted to the Protein Sequence Database, July 2000
A:Reference number: 225286
A:Accession: T51031
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-217 <SCH>
A:Cross-references: EMBL:AL389900; GSPDB:GN00116; NCSP:B15120.10
A:Experimental source: BAC clone B15120; strain OR74A
C:Genetics:
A:Gene: NCSP:B15120.10
A:Map position: 6

Query Match          18.5% ; Score 77.5 ; DB 2 ; Length 217 ;
Best Local Similarity 37.5% ; Pred. No. 0.9 ;
Matches 21 ; Conservative 5 ; Mismatches 21 ; Indels 9 ; Gaps 2 ;

OY      7 PRPAAVPYPLRMQGPCPAHVLISFLRPSMDIVSAFYSLPLADLSPTSVPISPSVGR 62

```

```

RESULT      9
Tl9673
hypothetical protein C33B4.3 - Caenorhabditis elegans
C|Species: Caenorhabditis elegans
C|Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
R|Accession: Tl9673
R|Coles, L.
Submitted to the EMBL Data Library, February 1995
A|Reference number: Z19160
A|Accession: Tl9673
A|Status: preliminary; translated from GB/EMBL/DDBJ
A|Molecule type: DNA
A|Residues: 1-1110 <MIL>
A|Cross-references: EMBL:Z48367; PIDN:CAA88324.1; GSPDB:GNO0020; CESP:C33B4.3
A|Experimental source: clone C33B4
C|Genetics:
A|Gene: CESP:C33B4.3
A|Map position: 2
A|Introns: 20/3; 110/3; 144/3; 341/1; 455/1; 546/1; 653/3; 934/2; 983/3; 1035/1

Query Match          18.3%; Score 76.5; DB 2; Length 1110;
Best Local Similarity 28.6%; Pred.No. 7.3;
Matches    24; Conservative   12; Mismatches    33; Indels     15; Gaps       4;

QY      3 HSLPRPAAVPPLRMQ-----PCGAHPVLSFLRPMDVSATFYL-LAPLSPTSVPS 56
        ||||| : | : | : | : | : | : | : | : | : | : | : | : | :
Db      744 HPSPRASSTPOPIOOOOSIIPPPPPPPhCEPT--MVHVEFTTPPSTSSVPppPPlP 801
               PVSVGRGDPD-----AHVAVN 73
Dd      802 PISSCAPPPPPPPPGGLMHVAAS 825

RESULT      10
A96826
TBKt4.10 [imported] - Arabidopsis thaliana
C|Species: Arabidopsis thaliana (mouse-ear cress)
C|Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
```

C:Accession: A96826
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
 Hansen, N.F.; Hughes, B.; Huizar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
 C.A.; Li, J.H.; Li, X.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Malt, R.; Marziani,
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
 Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719
 A:Accession: A96826
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-356 <STO>
 A:Cross-references: GB:AEO05173; NID:g4835761; PIDN:AA030228.1; GSPDB:GN00141
 C:Genetics:
 A:Gene: T8K14.10
 A:Map position: 1

Query Match 18.2%; Score 76; DB 2; Length 356;
 Best Local Similarity 34.7%; Pred. No. 2.2;
 Matches 26; Conservative 7; Mismatches 30; Indels 12; Gaps 3;

OY 9 PAAPVPLRMQGPAPHPVLSFLRPSMDLVSAF-----YSLPLAPLSPTSPISPV---VSQ 60
 DB 216 PCGVLCPPYSEGPSTPTGTSIPSPS-----SGFLPPIYVPPMAPSPSPVTSATYCAVAK 271
 OY 61 GRGPDPAHVAVNL 75
 DB 272 PSYDPPIIOEAMNFA 286

RESULT 11
 S48273
 probable transcription factor YBR108w - yeast (Saccharomyces cerevisiae)
 N:Alternate names: hypothetical protein YBR0901
 C:Species: Saccharomyces cerevisiae
 C:Date: 01-Aug-1995 #sequence_revision 01-Sep-1995 #text_change 21-Jul-2000
 C:Accession: S48273; S45976; S44688
 R:Manhaupt, G.; Stucka, R.; Ehme, S.; Vetter, I.; Felmann, H.
 Yeast 10, 1363-1381, 1994
 A:Title: Analysis of a 70 kb region on the right arm of yeast chromosome II.
 A:Reference number: S48255; MUID:95208357
 A:Accession: S48273
 A:Status: nucleic acid sequence not shown
 A:Molecule type: DNA
 A:Residues: 1-848 <MAN>
 A:Cross-references: EMBL:X78993; NID:g476045; PIDN:CAAS5611.1; PID:g476064
 R:Felmann, H.; Manhaupt, G.; Schwarzlose, C.; Vetter, I.
 submitted to the Protein Sequence Database, August 1994
 A:Reference number: S45927
 A:Accession: S45976
 A:Molecule type: DNA
 A:Residues: 1-848 <FE2>
 A:Cross-references: EMBL:Z35977; NID:g536378; PID:g536379; MIPS:YBR108w
 C:Genetics:
 A:Map position: 2R

Query Match 18.2%; Score 76; DB 2; Length 848;
 Best Local Similarity 34.3%; Pred. No. 6;
 Matches 23; Conservative 6; Mismatches 28; Indels 10; Gaps 3;

OY 13 PVPRLMQGPAPHPV-----ISFLRPSMDLVSA---FYSPLAPLSPTSPV---ISPVSGR 62
 DB 383 PVPVRMQGPAPHPV-----ISFLRPSMDLVSA---FYSPLAPLSPTSPV---ISPVSGR 62
 OY 63 GRPDPAN 69
 DB 443 LPPPTN 449

RESULT 12
 D49600
 genome-linked protein Vpg - soybean dwarf virus
 C:Species: soybean dwarf virus
 C:Date: 19-Mar-1997 #sequence_revision 19-Mar-1997 #text_change 20-Sep-1999
 C:Accession: D49600
 R:Rathjen, J.P.; Karageorgos, L.E.; Habbil, N.; Waterhouse, P.M.; Symons, R.H.
 Virology 198, 671-679, 1994
 A:Title: Soybean dwarf luteovirus contains the third variant genome type in the luteo
 A:Reference number: A49600; MUID:94120742
 A:Accession: D49600
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-189 <RAU>
 A:Cross-references: GB:L24049; NID:g436017; PIDN:AA17538.1; PID:g436021
 C:Superfamily: potato leaf roll virus genome-linked protein

Query Match 17.9%; Score 75; DB 2; Length 189;
 Best Local Similarity 31.1%; Pred. No. 1.4;
 Matches 19; Conservative 12; Mismatches 22; Indels 8; Gaps 2;

OY 2 THSLPRAAPVPL-----RMQGPAPHPVLSFLRPSMDLVSAF---FYSPLAPLSPTSPV 53
 DB 58 THSCQRTASMVVPPREVLSGRLYGNASHSLMEYSRPTWNIRSVSYSSSRPLPPQV 117
 OY 54 P 54
 DB 118 P 118

RESULT 13
 T48814
 hypothetical protein 15E6.220 [imported] - Neurospora crassa
 C:Species: Neurospora crassa
 C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 05-May-2000
 C:Accession: T48814
 R:Schultz, U.; Alqn, V.; Hohelsel, J.; Brandt, P.; Farmann, B.; Holland, R.; Nyakatu
 submitted to the Protein Sequence Database, April 2000
 A:Reference number: Z24541
 A:Accession: T48814
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-1952 <SCH>
 A:Cross-references: EMBL:AL353822; GSPDB:GN00112; NCSP:15E6.220
 A:Experimental source: cosmid contig 15E6; strain 74
 C:Genetics:
 A:Gene: NCSP:15E6.220
 A:Map position: 2
 A:introns: 281/3

Query Match 17.9%; Score 75; DB 2; Length 1952;
 Best Local Similarity 34.8%; Pred. No. 20;
 Matches 24; Conservative 8; Mismatches 31; Indels 6; Gaps 3;

OY 2 THSLPRAAPVPLRMQGPAPHPVLSFLRPSMDLVSAFYSPLAPLSPTSPV---ISPV 58
 DB 1781 TPQAVPRPSTALTPAQCPVSPAVS--GSGVPAASQAVAPAVSSTVPAAATVAPA 1838
 OY 59 S-VGRGPD 66
 DB 1839 STVAAAPT 1847

RESULT 14
 T13029
 beta-adaptin homolog F8L21.170 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cress)
 C:Date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 24-Nov-1999
 C:Accession: T13029

R:Bevan, M.; Peters, S.A.; van Staveren, M.; Dirkse, W.; Stiekema, W.; Bancroft, I.; Mew
submitted to the Protein Sequence Database, July 1999
A:Reference number: Z17587

A:Accession: T13029

A:Molecule type: DNA

A:Residues: 1-894 <BEV>

A:Cross-references: EMBL:AL096882; GSPDB:GN00062; ATSP:F8L21.170

A:Experimental source: cultivar Columbia; BAC clone F8L21

C:Genetics:

A:Gene: ATSP:F8L21.170

A:Map position: 4

A:Introns: 29/3; 95/3; 133/3; 217/2; 262/3; 322/3; 353/3; 424/2; 445/3; 505/3; 536/3; 66
C:Superfamily: beta-adaptin

Query Match 17.8%; Score 74.5; DB 2; Length 894;

Best Local Similarity 29.6%; Pred. No. 9.1;

Matches 24; Conservative 8; Mismatches 30; Indels 19; Gaps 3;

OY 9 PAAVVPRLMQPGPAHPVLSFLRPSMDLVSATYSLPLAPLSPTSVPISE-----PV 58

DB 619 PCNIFQPSGRQPAAPVPA-----PVPDLIGDLMGLDNAAIVPVDPITQSGPPLPVVPA 673

OY 59 SVGRGPPDPAHVAVNLISRYEG 79

DB 674 SSGQG---LQISAQLSRKDG 690

RESULT 15

T00273

hypothetical protein KIAA0595 - human (fragment)

C:Species: Homo sapiens (man)

C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000

C:Accession: T00273

R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Momura, N.; Ohara, O.

DNA Res. 5, 31-39, 1998

A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complet

A:Reference number: Z14086; MUID:98290545

A:Accession: T00273

A>Status: Preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-1520 <NAG>

A:Cross-references: EMBL:AB011167; NID:g3043713; PIDN:BAA25521.1; PID:g3043714

A:Experimental source: brain

C:Genetics:

A:Note: KIAA0595

Query Match 17.5%; Score 73; DB 2; Length 1520;

Best Local Similarity 35.1%; Pred. No. 23;

Matches 27; Conservative 3; Mismatches 25; Indels 22; Gaps 4;

OY 5 LIPRPA-----AVPYPLRMQPGPAHPVLSFLRPSMDLVSATYSLPLAPLSPTSVPISE 56

DB 710 LLARSPPVQSVSPAVPTPPSMSALPFPAGLGMP-----SLPPPLQPPSLPLS 761

OY 57 --PVSVGRGPPDPAHVA 71

DB 762 MGPVL---PDPTTHYA 774

Search completed: April 11, 2002, 09:30:07
Job time: 104 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:13 ; Search time 15.52 Seconds

(without alignments)
186.632 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418
1 GTHSLPRAAVPVPLRMQP.....VGRGPDPAHVAVNLSRYEG 79

Sequence: BLOSUM62
Gapop 10.0 , Gapext 0.5

Scoring table: 100059 seqs, 3664827 residues

Searched: Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : SwissProt_39*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	83.5	20.0	888	SM6B_HUMAN	Q9H3T3 homo sapien
2	81.5	19.5	1257	PGCN_RAT	P55067 ratius norv
3	80	19.1	428	FXB2_MOUSE	Q64733 mus musculu
4	80	19.1	1840	RPB1_ARATH	P18616 arabidopsis
5	76	18.2	848	YBVB_YEAST	P38266 saccharomyc
6	74.5	17.8	1255	PER2_HUMAN	Q15055 homo sapien
7	73.5	17.6	864	YC18_HUMAN	Q9ULX2 homo sapien
8	73.5	17.6	887	SM6B_RAT	O70141 ratius norv
9	72	17.2	440	CBP_MOUSE	P30098 homo sapien
10	71.5	17.1	2441	CBP_MOUSE	P45481 mus musculu
11	70.5	16.9	291	YD53_STYNY3	P42350 synechocyst
12	70.5	16.9	788	DPOC_HPBHE	P13846 heron hepat
13	70.5	16.9	1268	PGCN_MOUSE	P55066 mus musculu
14	70	16.7	2318	NTC3_MOUSE	Q61982 mus musculu
15	69.5	16.6	433	TEF5_CHICK	Q90701 gallus galli
16	69	16.5	265	CDX1_HUMAN	P47907 homo sapien
17	68.5	16.4	886	SM6B_MOUSE	O54951 mus musculu
18	68	16.3	703	MUS1_HUMAN	Q05823 homo sapien
19	68	16.3	1206	FM14_MOUSE	Q05833 mus musculu
20	68	16.3	1468	FMN1_MOUSE	Q05860 mus musculu
21	68	16.3	1752	RPB1_SCHPO	P36594 schizosach
22	68	16.3	1859	RPB1_CAEEL	P16356 caenorhabdi
23	67.5	16.1	123	VST1_HEYBU	P29325 hepatitis e
24	67.5	16.1	123	VST1_HEYMY	O04612 hepatitis e
25	67.5	16.1	213	AMEX_BOVIN	P02817 bos taurus
26	67.5	16.1	506	ARSA_MOUSE	P50428 mus musculu
27	67.5	16.1	667	SIX5_MOUSE	P70178 mus musculu
28	67.5	16.1	827	SOM6_MOUSE	P40645 mus musculu
29	67	16.0	360	A2H8_RABIT	P80191 cryocolagus
30	67	16.0	817	VRP1_YEAST	P37370 saccharomyc
31	67	16.0	1433	Y310_HUMAN	O15027 homo sapien
32	66.5	15.9	259	MSP8_EIMAC	P09125 elmeria ace
33	66.5	15.9	283	EXTN_SORBI	P24152 sorghum bic

34	66.5	15.9	435	1	TEF5_HUMAN	Q99594 homo sapien
35	66.5	15.9	1229	1	N121_HUMAN	O9Y2N3 homo sapien
36	66	15.8	534	1	APG_ARATH	P40602 arabidopsis
37	66	15.8	538	1	TF65_CHICK	P98152 gallus galli
38	66	15.8	868	1	NRG2_RAT	O35569 ratius norv
39	66	15.8	1447	1	DCG_HUMAN	P43146 homo sapien
40	66	15.8	1447	1	DCG_MOUSE	P70211 mus musculu
41	66	15.8	3149	1	TEGU_EBV	P03186 epstein-bar
42	65.5	15.7	357	1	GDF1_MOUSE	P20863 mus musculu
43	65.5	15.7	1257	1	PER2_MOUSE	O54943 mus musculu
44	65.5	15.7	2805	1	MAPA_HUMAN	P78559 homo sapien
45	65	15.6	582	1	MNT_HUMAN	Q99583 homo sapien

ALIGNMENTS

RESULT	ID	SM6B_HUMAN	STANDARD	PRT	888 AA.
AC	O9H3T3	20-AUG-2001 (Rel. 40, Created)			
DT	20-AUG-2001 (Rel. 40, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	SEMAPHORIN 6B PRECURSOR (SEMAPHORIN Z) (SEMA Z).				
GN	SEMA6B OR SEMAZ.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	(1)				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Brain;				
RA	Kimura T., Ishida H.;				
RL	Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.				
CC	- FUNCTION: MAY PLAY A ROLE IN BOTH PERIPHERAL AND CENTRAL NERVOUS				
CC	SYSTEM DEVELOPMENT (BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				
CC	- SIMILARITY: BELONGS TO THE SEMAPHORIN FAMILY.				
CC	- SIMILARITY: CONTAINS 1 SEMA DOMAIN.				
CC	-----				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; AB022433; BAB20659.1; .				
DR	InterPro; IPR003659; PSI.				
DR	InterPro; IPR001627; Sema.				
DR	InterPro; IPR000737; Squash.				
DR	Pfam; PF01403; Sema; 1.				
DR	SMART; SM00423; PSI; 1.				
DR	SMART; SM00286; PTI; 1.				
KW	Signal. Transmembrane; Multigene family; Neurogenesis; Glycoprotein;				
KW	Developmental protein.				
FT	SIGNAL	1	25		POTENTIAL.
FT	CHAIN	26	888		SEMAPHORIN 6B.
FT	DOMAIN	26	603		EXTRACELLULAR (POTENTIAL).
FT	TRANSEM	604	624		POTENTIAL.
FT	DOMAIN	625	888		CYTOPLASMIC (POTENTIAL).
FT	DOMAIN	238	547		SEMA.
FT	DOMAIN	661	674		POLY-GLY.
FT	DOMAIN	750	753		POLY-LEU.
FT	CARBOHYD	74	74		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	155	155		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	167	167		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	291	291		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	386	386		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	441	441		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	462	462		N-LINKED (GLCNAC. . .) (POTENTIAL).

DB 670 SVPEQAVRPVSFC-AEDPE 688

RESULT 3
ID FXB2_MOUSE STANDARD: PRT: 428 AA.

AC 064733:
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE FORKHEAD BOX PROTEIN B2 (TRANSCRIPTION FACTOR FKX-4).
GN FOXB2 OR FKX4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97014266; PubMed=8611101;
RA Kaestner K.H., Schuetz G., Monaghan A.P.;
RT "Expression of the winged helix genes Fkh-4 and Fkh-5 defines domains
in the central nervous system.";
RL Mech. Dev. 55:221-230(1996).
RN [2]
RP SEQUENCE OF 4-114 FROM N.A.
RX STRAIN=129;
RA MEDLINE=93361500; PubMed=7689224;
RA Kaestner K.H., Lee K.H., Schloendorff J., Hiemisch H.,
RA Monaghan A.P., Schuetz G.;
RT "Six members of the mouse forkhead gene family are developmentally
regulated.";
RN [2]
RL Proc. Natl. Acad. Sci. U.S.A. 90:7628-7631(1993).
CC -1 SUBCELLULAR LOCATION: NUCLEAR.
CC -1 DEVELOPMENTAL STAGE: EXPRESSED DURING EMBRYOGENESIS.
CC -1 SIMILARITY: CONTAINS 1 FORK-HEAD DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

CC
DR EMBL: X92591; CA63335.1; -
DR EMBL: X71942; CA50744.1; -
DR TRANSPAC: T02442; -
DR MGD: MGI:1347468; Foxd2.
DR InterPro: IPR001766; Fork_head.
DR Pfam: PF00250; Fork_head, 1.
DR PRINTS: PR00053; FORKHEAD.
DR SMART: SM00339; Fork_1.
DR PROSITE: PS00657; FORK_HEAD_1; 1.
DR PROSITE: PS00658; FORK_HEAD_2; 1.
DR PROSITE: PS00639; FORK_HEAD_3; 1.
KM DNA-binding; Nuclear Protein; Transcription regulation.
FT DNAS_BIND 12 103
FT DNAS_BIND 139 153
FT DNAS_BIND 156 162
FT DNAS_BIND 163 172
FT DNAS_BIND 217 231
FT DNAS_BIND 249 258
FT DNAS_BIND 321 330
FT DNAS_BIND 396 399
SQ SEQUENCE 428 AA; 45170 MW; DB8ABEF1E94AB10 CRC64;

Query Match 19.1%; Score 80; DB 1; Length 428;
Best Local Similarity 34.8%; Pred. No. 0.66;
Matches 23; Conservative 10; Mismatches 25; Indels 8; Gaps 2;
OY 3 HSLPRPAVAVPRLRMQCGAPVLSFLRPSMDLVSAFYSLP-----LAPLSPFVSVIS 56

DB 350 HSNQSLPAVAVPIK--PTPALPEVTTLPALSVPTASQQLPASTVCAAAASTAPLLE 407

OY 57 PVSAGR 62
DB 408 PTAAGR 413

RESULT 4
ID RPB1_ARATH STANDARD: PRT: 1840 AA.

AC P18616; P31635; O9S2S8;
DT 01-NOV-1990 (Rel. 16, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE DNA-DIRECTED RNA POLYMERASE II LARGEST SUBUNIT (RC 2.7.7.6).
GN RPB205 OR RPII OR RPB1 OR AT4G35800 OR F4B14.70.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC Eustroids II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=CV, COLUMBIA;
RX MEDLINE=91080867; PubMed=2259344;
RA Nawrath C., Schell J., Koncz C.;
RT "Homologous domains of the largest subunit of eucaryotic RNA
polymerase II are conserved in plants.";
RN [2]
RL Mol. Genet. 223:65-75(1990).

RP SEQUENCE FROM N.A.
RX STRAIN=CV, COLUMBIA;
RX MEDLINE=91355869; PubMed=2103447;
RA Dietrich M.A., Prenger J.P., Guilfoyle T.J.;
RT "Analysis of the genes encoding the largest subunit of RNA polymerase
II in Arabidopsis and soybean.";
RN [2]
RL Plant Mol. Biol. 15:207-223(1990).

CC [3]
CC SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RX MEDLINE=2003488; PubMed=10617198;
RA Mayer K.F.X., Schueller C., Wambutt R., Murphy G., Volckaert G.,
RA Pohl T., Duesterhoeft A., Stiekema W., Entian K.-D., Terryn N.,
RA Harris B., Anstorge W., Brandt P., Grivell L.A., Rieger M.,
RA Weichselgartner M., de Simone V., Obermaier B., Maché R., Mueller M.,
RA Kreis M., Delseny M., Puigdomenech P., Watson M., Schmidthein T.,
RA Reichert B., Portetelle D., Perez-Alonso M., Boutry M., Bancroft I.,
RA Vos P., Hohenseil J., Zimmermann W., Wedler H., Ridley P.,
RA Langham S.-A., McCullagh B., Bilham L., Robben J.,
RA Van der Schueren J., Grymonprez B., Chuang Y.-J., Vandenbussche F.,
RA Braeken M., Weltjens I., Voet M., Bastiaens I., Aert R., Defoor E.,
RA Weltzenerger T., Bothe G., Ramsperger U., Hilbert H., Braun M.,
RA Holzer E., Brandt A., Peters S., van Staaveren M., Dirkes W.,
RA Moolman P., Klein Lankhorst R., Rose M., Haut J., Koetter P.,
RA Berneiser S., Hempel S., Feltpausch M., Lamberth S., Van den Daele H.,
RA De Keyser A., Buysshaert C., Gielen J., Villarroel R., De Clercq R.,
RA Van Montagu M., Rogers J., Cronin A., Quail M., Bray-Allen S.,
RA Clark L., Doggett J., Hall S., Kay M., Leonard N., McIay K., Mayes R.,
RA Bertlett A., Randream M.-A., Lyné M., Benes V., Rechmann S.,
RA Borkova D., Blocker H., Scharfe M., Grimm M., Loehert T.-H.,
RA Dose S., de Haan M., Maarse A.C., Schaefer M., Mueller-Auer S.,
RA Gabel C., Fuchs M., Fartmann B., Grandtath K., Dauner D., Hezli A.,
RA Neumann S., Argitlou A., Vitale D., Liguori R., Piravandi E.,
RA Massenat O., Quigley F., Clapaud G., Muendlein A., Felber R.,
RA Schnabel S., Hiller R., Schmidt W., Lecharry A., Aubourg S.,
RA Chefor F., Cooke R., Berger C., Monfort A., Casacuberta E.,
RA Gibbons T., Weber N., Vandenbol M., Bagues M., Terol J., Torres A.,
RA Perez-Perez A., Purnelle B., Bent E., Johnson S., Tacon D., Jesse T.,
RA Heijnen L., Schwarz S., Scholler P., Heber S., Frances P., Bietke C.,
RA Frisman D., Haase D., Lemcke K., Mewes H.-W., Stocker S.,
RA Zaccaria P., Beyan M., Wilson R.K., de la Bastide M., Habermann K.,
RA Parnell L., Dedhia N., Gnoj L., Schutz K., Huang E., Spiegel L.,

	RA	Stehonk M., Murray J., Sheet P., Cordes M., Abu-Threideh J.,
	RA	Raichlen D., Kalicki J., Graves T., Harmon G., Edwards J.,
	RA	Lacaille P., Courtney L., Cloud J., Abbott A., Scott K., Johnson D.,
	RA	Mix P., Bentley D., Fulton B., Miller N., Greco T., Kemp K.,
	RA	Krimer J., Patton L., Marks E., Dalton M., Pepin K., Hillier L.,
	RA	Nelson J., Speith J., Ryan E., Andrews S., Gelsel C., Layman D.,
	RA	Du H., Ali J., Berghoff A., Jones K., Drone K., Cotton M., Joshi C.,
	RA	Antonouli B., Zidane M., Strong C., Sun H., Lamar B., Yordan C.,
	RA	Ma P., Zhong Y., Preston R., Vil D., Shekher M., Mateo A., Shah R.,
	RA	Svaby I.K., O'Shaughnessy A., Rodriguez M., Hoffman J., Till S.,
	RA	Grenat S., Shohdy N., Hasegawa A., Hameed A., Lodhi M., Johnson A.,
	RA	Chen E., Maria M., Martienssen R., McCombie W.R.
	RT	"Sequence and analysis of chromosome 4 of the plant Arabidopsis
	RT	thaliana."
	RL	Nature 402:769-777(1999).
	CC	- FUNCTION: DNA-DEPENDENT RNA POLYMERASE CATALYZES THE TRANSCRIPTION
	CC	OF DNA INTO RNA USING THE FOUR RIBONUCLEOSIDE TRIPHOSPHATES AS
	CC	SUBSTRATES.
	CC	- CATALYTIC ACTIVITY: N NUCLEOSIDE TRIPHOSPHATE - N PYROPHOSPHATE +
	CC	RNA(N).
	CC	- SUBCELLULAR LOCATION: NUCLEAR.
	CC	- 1-PJM: THE TANDEM 7 RESIDUES REPEATS CAN BE HIGHLY PHOSPHORYLATED.
	CC	THE PHOSPHORYLATION ACTIVATES POL2.
	CC	- MISCELLANEOUS: THREE DISTINCT ZINC-CONTAINING RNA POLYMERASES WERE
	CC	FOUND IN EUKARYOTIC NUCLEI: POLYMERASE I FOR THE RIBOSOMAL RNA
	CC	PRECURSOR, POLYMERASE II FOR THE MRNA PRECURSOR, AND POLYMERASE
	CC	III FOR 5S AND TRNA GENES.
	CC	- SIMILARITY: BELONGS TO THE RNA POLYMERASE BETA' CHAIN FAMILY.
	CC	-----
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
	CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
	CC	the European Bioinformatics Institute. There are no restrictions on its
	CC	use by non-profit institutions as long as its content is in no way
	CC	modified and this statement is not removed. Usage by and for commercial
	CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/sib-ch/).
	CC	or send an email to license@sib-sib.ch .
	CC	-----
	DR	EMBL; X52954; CA337130.1; -
	DR	-EMBL; X52954; CA337130.1; -
	DR	EMBL; AL031986; CA21466.2; -
	DR	EMBL; AL031986; CA21466.2; -
	DR	PIR; S12071; JDM01.
	DR	PIR; S11960; JDM02.
	DR	InterPro; IPR002865; P-rich_extensn.
	DR	InterPro; IPR000684; RNA_pol_II_repeat.
	DR	InterPro; IPR000722; RNA_pol_A.
	DR	InterPro; IPR002879; RNA_pol_A2.
	DR	Pfam; PF00623; RNA.pol.A.1.
	DR	Pfam; PF01854; RNA.pol.A2.1.
	DR	PROSITE; PS00115; RNA_POL_II_REPEAT; 23.
	DR	transferase; DNA-directed RNA polymerase; Transcription; Zinc; Repeat;
	DR	DNA-binding; Nuclear protein; Nucleic acid binding; Zinc-finger.
	KW	DNA-binding; Nuclear protein; Nucleic acid binding; Zinc-finger.
	FT	ZN_FING 66 79
	FT	DNA_BIND 327 398
	FT	DOMAIN 786 796
	FT	DOMAIN 1530 1819
	FT	CONFLICT 117 124
	FT	CONFLICT 193 197
	FT	CONFLICT 298 298
	FT	CONFLICT 303 303
	FT	CONFLICT 401 417
	FT	CONFLICT 428 428
	FT	CONFLICT 446 446
	FT	CONFLICT 739 739
	FT	CONFLICT 1062 1062
	FT	CONFLICT 1089 1089
	FT	CONFLICT 1720 1720
	SO	SEQUENCE 1840 AA; 204688 MW; 8453621AD945C1B6 CRC64;

```

Best Local Similarity 36.4%; Pred. No. 3.1;
Matches 28; Conservative 3; Mismatches 36; Indels 10; Gaps 3;

QY 7 PRPAAPVPLRMQP-----GPAHPVLSTFRPSMDLVSAFYSLPLAPLSPTSVIPSVYGR 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1736 PSIAVSPSNARLSPASPYSPSPVNSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSGA 1791

QY 63 GPD--PDHVAVNLNLSRY 77
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 1792 SPDYSPSAGYSPTLPGY 1808

RESULT 5
ID YBV8_YEAST STANDARD; PRT; 848 AA.
AC P38266;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-DEC-1994 (Rel. 37, Last annotation update)
DE HYCOTHEETICAL 92.8 KDA PROTEIN IN PH088-CMD1 INTERGENIC REGION.
GN YBR108W OR YBR0901.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX NCBI_TaxID=4932;
[1]
RP SEQUENCE FROM N.A.
RN STRAIN=S288C;
RX MEDLINE=95208357; PubMed=7900426;
RT Mannhaupt G., Stucke R., Ehnlé S., Vetter I., Feldmann H.;
RL "Analysis of a 70 kb region on the right arm of yeast chromosome II.";
      Yeast 10:1363-1381(1994).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC CC the European Bioinformatics Institute. There are no restrictions on its
CC CC use by non-profit institutions as long as its content is in no way
CC CC modified and this statement is not removed. Usage by and for commercial
CC CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC CC or send an email to license@isb-sib.ch).
CC CC -----
CC CC EMBL: X78993; CAA55611.1; -
DR DR EMBL: Z35977; CAA85063.1; -
DR DR PIR: S44688; S44688.
DR DR SGD: S0000312; YBR108W.
RW KW Hypothetical protein.
SQ SEQUENCE 848 AA; 92762 MW; F33D371369FBAF97 CRC64;

Query Match 18.2%; Score 76; DB 1; Length 848;
Best Local Similarity 34.3%; Pred. No. 3.4;
Matches 23; Conservative 6; Mismatches 28; Indels 10; Gaps 3;

QY 13 PVPFLMOPGPAHPV-----LSFLRPSMDLVSA---FYSPLAPLSPTSVP---ISPSYGR 62
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 363 PVPVAMQGPQPPMQGNGINYPLEPSLDSTGSRPHFVYFPDPDAPAPRKPIDIPVYSS 442

QY 63 GPDPAH 69
      | | |
DB 443 LPPPTH 449

RESULT 6
ID PER2_HUMAN STANDARD; PRT; 1255 AA.
AC O15055;
DT 15-JUL-1999 (Rel. 38, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PERIOD CIRCADIAN PROTEIN 2.
GN PER2 OR KIAA0347.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

```

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain:
RX MEDLINE=97349984; PubMed=9205841;
RA Nagase T., Ishikawa K.-I., Nakajima D., Ohira M., Seki N.,
Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. VII.
The complete sequences of 100 new cDNA clones from brain which can
code for large proteins in vitro.";
RL DNA Res. 4:141-150(1997).
RN [2]
RP REVISIONS TO C-TERMINUS.
RC TISSUE-Brain:
RA Nagase T., Ishikawa K.-I., Seki N., Nakajima D., Ohira M.,
Miyajima N., Kotani H., Nomura N., Ohara O.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP TISSUE EXPRESSION.
RX MEDLINE=98087121; PubMed=9427249;
RA Shearman L.P., Zylka M.J., Weaver D.R., Kolakowski L.F. Jr.,
Reppert S.M.;
RT "Two period homologs: circadian expression and photic regulation in
the suprachiasmatic nuclei.";
RL Neuron 19:1261-1269(1997).
CC -1- FUNCTION: CIRCADIAN REGULATOR THAT MAY ACT AS A TRANSCRIPTION
FACTOR. BEHAVES AS A NEGATIVE ELEMENT IN CIRCADIAN TRANSCRIPTIONAL
LOOP. DOES NOT APPEAR TO BIND DNA, SUGGESTING INDIRECT
TRANSCRIPTIONAL INHIBITION. EXPRESSION OSCILLATES WITH A 24 HR
RHYTHM IN THE SUPRACHIASMATIC NUCLEUS (SCN) AND THE WHOLE EYES.
OSCILLATIONS ARE MAINTAINED UNDER CONSTANT DARKNESS AND ARE
RESPONSIVE TO CHANGES OF THE LIGHT/DARK CYCLES. THERE IS A 4 HOUR
TIME DELAY BETWEEN PER1 AND PER2 OSCILLATIONS. THE EXPRESSION
RHYTHMS APPEAR TO ORIGINATE FROM RETINA (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
CC -1- TISSUE SPECIFICITY: WIDELY EXPRESSED. FOUND IN HEART, BRAIN,
PLACENTA, LUNG, LIVER, SKELETAL MUSCLE, KIDNEY AND PANCREAS.
CC -1- INDUCTION: BY LIGHT (BY SIMILARITY).
CC -1- SIMILARITY: CONTAINS 1 PAS (PER-ARNT-SIM) DIMERIZATION DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE BASIC HELIX-LOOP-HELIX (BHLH) FAMILY OF
TRANSCRIPTION FACTORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AB002345; BAA20804.2; ALT_INIT.
DR MIM; 603426; .
DR InterPro; IPR001610; PAC.
DR InterPro; IPR000014; PAS.
DR SMART; SMO0086; PAC; 1.
DR SMART; SMO0091; PAS; 2.
KW Transcription regulation; Nuclear protein; Repeat; Biological rhythms.
FT DOMAIN 109 146 HELIX-LOOP-HELIX MOTIF (BY SIMILARITY).
FT REPEAT 182 248 PAS-1.
FT REPEAT 322 386 PAS-2.
FT DOMAIN 398 438 PAC MOTIF.
FT DOMAIN 510 513 POLY-ARG.
FT DOMAIN 789 806 NUCLEAR LOCALIZATION SIGNAL.
FT DOMAIN 842 979 PRO-RICH.
SQ SEQUENCE 1255 AA; 136579 MW; 2AEF2C6BD4B6CB0 CRC64;

Query Match 17.8%; Score 74.5; DB 1; Length 1255;
Best Local Similarity 25.6%; Pred. No. 7.2;
Matches 30; Conservative 10; Mismatches 26; Indels 51; Gaps 5;

OY 1 GTHSLPRP-----AAVPLRLMO-----PGPAHPVLSPFLRPSMDLVSAFYSLPL 45

DB 863 GTVAAPAPAPHASFVPAVAVDLOHQFAVOPPPFPAPLAPVMAFWLPSYSPSGCPNLPQ 922
OY 46 A-PLSP-----TSVP-----ISVSVGRGDP 66
DB 923 AFFPSQPOFSPSHPTLTSEMASASOPEFSPRTSIPQPCACPATRATPPSAMGRASP 979
RESULT 7
YC18_HUMAN STANDARD; PRT; 864 AA.
ID YC18_HUMAN
AC Q90LK2;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN KIAA1218 (FRAGMENT).
GN KIAA1218.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-Brain:
RX MEDLINE=20039619; PubMed=10574462;
RA Nagase T., Ishikawa K.-I., Kikuno R., Hirose M., Nomura N.,
Ohara O.;
RT "Prediction of the coding sequences of unidentified human genes. XV.
The complete sequences of 100 new cDNA clones from brain which code
for large proteins in vitro.";
RL DNA Res. 6:337-345(1999).
CC -1- SIMILARITY: BELONGS TO THE ATAXIN 7 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch)
CC -----
DR EMBL; AB033044; BAA6532.1; .
KW Hypothetical protein.
FT NON_TER 1
FT DOMAIN 223 226 POLY-THR.
FT DOMAIN 652 664 POLY-SER.
FT DOMAIN 786 790 POLY-SER.
SQ SEQUENCE 864 AA; 92138 MW; DE5CB8130E48DA23 CRC64;

Query Match 17.6%; Score 73.5; DB 1; Length 864;
Best Local Similarity 35.5%; Pred. No. 6.1;
Matches 33; Conservative 4; Mismatches 21; Indels 35; Gaps 7;

OY 4 SLPRPA---VPLRL-MOPGAHPVLSPFLRPSMDLVSAFYSLPLAFLSP-----T 51
DB 517 SPLSPAAHHTTPVPASVLP-----FSNP-----SAVY-LPSAPITSSRLTSSITMT 562
OY 52 SVPISPVSVGRGDPDA-----HYAVNLVS 75
DB 563 SAMLSNAFVTSPPSALMSHTTAFPHVAATLS 595
RESULT 8
SM6B_RAT STANDARD; PRT; 887 AA.
ID SM6B_RAT
AC O70141;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SEMAPHORIN 6B PRECURSOR (SEMAPHORIN Z) (SEMA Z).
GN SEMA6B.
OS Rattus norvegicus (Rat).

Query Match 17.2% Score 72; DB 1; Length 440;
 Best Local Similarity 45.2%; Pred.No. 4.2;
 Matches 19; Conservative 5; Mismatches 10; Indels 8; Gaps 3;

OY 2 THSLPRAAVPVPLRMQGPAPHLVSLRPSNDLVSAFYSL 43
 DB 32 TASPSPAPSWTPSP---RPGPAH---FLQPPMAV--ALMSL 65

RESULT 10
 CBP_MOUSE
 ID CBP_MOUSE STANDARD: PRT; 2441 AA.

AC P45481:
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-FEB-1996 (rel. 33, Last sequence update)
 DT 15-JUL-1998 (rel. 36, Last annotation update)
 DE CREB-BINDING PROTEIN.
 GN CREBBP OR CBP.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 OX NCBI_TaxId=10090;
 RN [1]
 RP SEQUENCE FROM N.A.

RC TISSUE=Brain;
 RX MEDLINE=94019866; PubMed=8413673;
 RA Christia J.C., Kwok R.P.S., Lamb N., Haglwaara M., Montminy M.R.,
 RA Goodman R.H.;
 RT "Phosphorylated CREB binds specifically to the nuclear protein CBP.";
 RL Nature 365:855-859(1993).

CC - FUNCTION: MEDIATES CAMP-GENE REGULATION BY BINDING SPECIFICALLY TO
 PHOSPHORYLATED CREB PROTEIN. ACTING AS A COACTIVATOR, CBP AUGMENTS
 THE ACTIVITY OF PHOSPHORYLATED CREB TO ACTIVATE TRANSCRIPTION OF
 CAMP-RESPONSIVE GENES.

CC - SUBCELLULAR LOCATION: NUCLEAR.
 CC - SIMILARITY: CONTAINS 1 BROMODOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; S66385; AAB28651.1; -;
 DR TRANSFAC; T01318; -;
 DR MGD; MGI:1098280; Crebbp.
 DR InterPro; IPR001487; Bromodomain.
 DR InterPro; IPR003101; KIX.
 DR InterPro; IPR000197; TAF_finger.
 DR InterPro; IPR000433; Znf_ZZ.
 DR Pfam; PF00439; bromodomain; 1.
 DR Pfam; PF02172; KIX; 1.
 DR Pfam; PF02135; zf-TAF; 2.
 DR Pfam; PF00569; ZF; 1.
 DR PRINTS; PR00503; BROMODOMAIN.
 DR SMART; SM00291; ZNF_ZZ; 1.
 DR SMART; SM00291; BROMO; 1.
 DR PROSITE; PS00633; BROMODOMAIN_1; 1.
 DR PROSITE; PS50014; BROMODOMAIN_2; 1.
 KW Transcription regulation; Nuclear protein; Activator; Bromodomain.
 FT DOMAIN 1104 1176 BROMODOMAIN.
 FT DOMAIN 1062 1065 POLY-GLU.
 FT DOMAIN 1556 1563 POLY-GLU.
 FT DOMAIN 1944 1949 POLY-PRO.
 FT DOMAIN 1968 1971 POLY-GLN.
 FT DOMAIN 2082 2086 POLY-GLN.
 FT DOMAIN 2200 2216 POLY-GLN.
 FT DOMAIN 2296 2299 POLY-GLN.
 FT SEQUENCE 2441 AA; 265474 MW; 0ABB028C3112FA19 CRC64;

Query Match 17.1% Score 71.5; DB 1; Length 2441;
 Best Local Similarity 36.5%; Pred.No. 28;
 Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

OY 4 SLPLPRAAVPVPLRMQGPAPHLVSLRPSNDLVSAFYSLPAPLSPVSPISVSGRG 63
 DB 843 SQLPCRPVQSLPHRPAPASNAAGN-PSLQHPTAPGHTPPQAPAPQ-PSTPVSSGQT 899

OY 64 PDP 66
 DB 900 PTP 902

RESULT 11
 YD53_SYNY3
 ID YD53_SYNY3 STANDARD: PRT; 291 AA.

AC P42350;
 DT 01-NOV-1995 (rel. 32, Created)
 DT 01-NOV-1995 (rel. 32, Last sequence update)
 DT 20-AUG-2001 (rel. 40, Last annotation update)
 DE HYPOTHETICAL 31.3 KDA PROTEIN SLR1353.
 GN SLR1353.
 OS Synechocystis sp. (strain PCC 6803).
 OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI_TaxId=1148;
 RN [1]
 RP SEQUENCE FROM N.A.

RX MEDLINE=9322488; PubMed=8467083;
 RA Malakhov M.P., Wada H., Los D.A., Sakamoto T., Murata N.;
 RT "Structure of a cyanobacterial gene encoding the 50S ribosomal
 protein L9.";
 RL Plant Mol. Biol. 21:913-918(1993).

RN [2]
 RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;
 RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
 RA Miyajima N., Hirosewa M., Sugiyama M., Sasamoto S., Kimura T.,
 RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Nairo K., Okumura S.,
 RA Shimo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
 RA Tabata S.;
 RT "Sequence analysis of the genome of the unicellular cyanobacterium
 Synechocystis sp. strain PCC6803. II. Sequence determination of the
 entire genome and assignment of potential protein-coding regions.";
 RL DNA Res. 3:109-136(1996).

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----
 DR EMBL; D10716; BAA38818.1; -;
 DR EMBL; D90912; BAA18174.1; -;
 KW Hypothetical protein; Complete proteome.
 SO SEQUENCE 291 AA; 31283 MW; 5AB7E3DD03C36390 CRC64;

Query Match 16.9% Score 70.5; DB 1; Length 291;
 Best Local Similarity 28.8%; Pred.No. 3.8;
 Matches 23; Conservative 16; Mismatches 30; Indels 11; Gaps 5;

OY 4 SLPLPRAAVPVPLRMQGPAPHLVSLRPSNDLVSAFYSLPAPLSPVSPISVSGRG 54
 DB 145 AIAPEVTLTPAPISPPSP-DEVLSLEETPPPPAVNNSFNQPESSAPIDSELQIDFAPR 203

OY 55 ISPVSYGRGPD-PDAHVAVN 73
 DB 204 ELPLAVEARPDSPEDMAVS 223

```

RESULT 12
DPOOL_HPBHE STANDARD; PRT; 788 AA.
AC P13846;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-JAN-1990 (Rel. 13, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE P PROTEIN [INCLUDES: DNA-DIRECTED DNA POLYMERASE (EC 2.7.7.7); RNA-
DIRECTED DNA POLYMERASE (EC 2.7.7.49); RIBONUCLEASE H (EC 3.1.26.4)].
GN
OS Heron hepatitis b virus.
OC Viruses; Retroid viruses; Hepadnaviridae; Avihepadnavirus.
OX NCBI_TaxID=28300;
RN [1]
RX MEDLINE=68333160; PubMed=3418788;
RA Sprengel R., Kuleta E.F., Will H.;
RT "Isolation and characterization of a hepatitis B virus endemic in
herons."
RL J. Virol. 62:3832-3839(1988).
CC -1- CATALYTIC ACTIVITY: N DEOXYNUCLEOSIDE TRIPHOSPHATE -
N PROPHOSPHATE + DNA(N).
CC -1- CATALYTIC ACTIVITY: ENDONUCLEOLYTIC CLEAVAGE TO 5'-PHOSPHO-
MONESTER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M22056; AAA5738.1;
DR PIR: A30082; JDVLMH.
DR InterPro: IPR001462; DNAPol_viral_C.
DR InterPro: IPR000201; DNAPol_viral_N.
DR InterPro: IPR000477; RVTse.
DR Pfam: PF00336; DNA_pol_viral_C_1.
DR Pfam: PF00242; DNA_pol_viral_N_1.
DR Pfam: PF00078; Tvc_1.
DR ProDom: PD000814; DNAPol_viral_C_1.
DR Transfaser: RNA-directed DNA polymerase; DNA-directed DNA polymerase;
KW Hydrolyase; Nuclease; Endonuclease; DNA replication; DNA-binding.
SO SEQUENCE 788 AA; 90070 MW; FB44F38F75EAD644 CRC64;

```

```

Query Match 16.9%; Score 70.5; DB 1; Length 788;
Best Local Similarity 40.0%; Pred. No. 11;
Matches 18; Conservative 4; Mismatches 12; Indels 11; Gaps 1;

```

```

QY 31 RPSWDVSAFYSLPLAPLSTSPIS-----PVSVGKCP 64
DB 444 RSLDLSQAFYHPLAPASSRLAVSDGKQYVYERKAPMGVGLSP 488

```

```

RESULT 13
PGCN_MOUSE STANDARD; PRT; 1268 AA.
AC P55066;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE NEUROCAN CORE PROTEIN PRECURSOR.
GN CSPG3 OR NCAN.
OS Mus musculus (Mouse).
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BA1B/C; TISSUE=Brain;

```

```

RX MEDLINE=96039250; PubMed=7490074;
RA Rauch U., Grimpe B., Kube G., Arnold-Ammer I., Beter D.,
RA Faessler R.;
RT "Structure and chromosomal localization of the mouse neurocan gene.";
RL Genomics 28:405-410(1995).
CC -1- FUNCTION: MAY MODULATE NEURONAL ADHESION AND NEURITE GROWTH DURING
CC DEVELOPMENT BY BINDING TO NEURAL CELL ADHESION MOLECULES (NC-CAM
CC AND N-CAM). CHONDROITIN SULFATE PROTEOGLYCAN; BINDS TO HYALURONIC
CC ACID.
CC -1- TISSUE SPECIFICITY: BRAIN.
CC -1- SIMILARITY: CONTAINS 1 IMMUNOGLOBULIN-LIKE V-TYPE DOMAIN.
CC -1- SIMILARITY: CONTAINS 2 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 2 LINK DOMAINS.
CC -1- SIMILARITY: CONTAINS 1 C-TYPE LECTIN FAMILY DOMAIN.
CC -1- SIMILARITY: CONTAINS 1 SUSHI (SCR) DOMAIN.
CC -1- SIMILARITY: BELONGS TO THE AGGRECAN/VERSICAN PROTEOGLYCAN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: X84727; CAA59216.1;
DR HSSP: P00740; IIXA.
DR MGD: MGI:104694; Cspg3.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001841; EGF_Ca.
DR InterPro: IPR003599; Ig.
DR InterPro: IPR003006; Ig_MHC.
DR InterPro: IPR000538; Link.
DR InterPro: IPR000436; Sush1_SCR_CCP.
DR InterPro: IPR001304; Lectin_C.
DR Pfam: PF00008; EGF_2.
DR Pfam: PF00047; Ig_1.
DR Pfam: PF00059; lectin_C_1.
DR Pfam: PF00084; sush1_1.
DR Pfam: PF00193; Xlink_2.
DR ProDom: PD000918; Link_2.
DR SMART: SM00032; CCP_1.
DR SMART: SM00034; CLECT_1.
DR SMART: SM00179; EGF_Ca_1.
DR SMART: SM00001; EGF_Like_1.
DR SMART: SM00409; IG_1.
DR SMART: SM00445; LINK_2.
DR PROSITE: PS00010; ASX_HYDROXYL; 1.
DR PROSITE: PS00022; EGF_1; 3.
DR PROSITE: PS00186; EGF_2; 1.
DR PROSITE: PS01187; EGF_Ca_1.
DR PROSITE: PS01241; LINK_2.
DR PROSITE: PS00615; C-TYPE-LECTIN_1; 1.
DR PROSITE: PS50041; C-TYPE-LECTIN_2; 1.
KW Glycoprotein; Hyaluronic acid; Proteoglycan; Immunoglobulin domain;
KW EGF-like domain; Calcium; Repeat; Lectin; Sush1; Signal.
FT SIGNAL 1 22
FT CHAIN 23 1268
FT DOMAIN 51 146
FT DOMAIN 176 253
FT DOMAIN 274 355
FT DOMAIN 960 996
FT DOMAIN 998 1034
FT DOMAIN 1036 1165
FT DOMAIN 1166 1224
FT DISULFID 58 139
FT DISULFID 181 252
FT DISULFID 205 226
FT DISULFID 279 354
FT DISULFID 303 324
FT DISULFID 964 975
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.
BY SIMILARITY.

```

Query Match 16.9% Score 70.5: DB 1: Length 1268:
Best Local Similarity 26.1%: Pred. No. 18;
Matches 30: Conservative 11; Mismatches 19; Indels 55; Gaps 7;

DB 604 LPR-PAAVVPLRMOPGPAH-----PVLSEFLR-----PSMDLVSA 39
ID 061982; STANDARD: PRT; 2318 AA.

QY 40 FYSLEPLAPLS--PTSPV---ISPVSVG-----RGPPDDA 68
DB 661 MTPVPLSPASPLSPWPEQAVRPSLGAEDLETPQTTIAPVENSHPDADS 715

RESULT 14
NTC3_MOUSE
AC 061982;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE "NEUROGENIC LOCUS NOTCH 3 PROTEIN."
GN NOTCH3.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ICR X SWISS WEBSTER;
RX MEDLINE=95001556; PubMed=7918097;
RA Lardelli M., Dalestrand J., Lendahl U.;
RT "The novel Notch homologue mouse Notch 3 lacks specific epidermal
RT growth factor-repeats and is expressed in proliferating
RT neuroepithelium".
RL Mech. Dev. 46:123-136(1994).
CC -1- FUNCTION: NOTCH 1, 2 AND 3 PLAY A COMBINATIONAL ROLE DURING
CC VARIOUS CELL FATE DECISIONS AND MORPHOLOGICAL MOVEMENTS IN THE
CC DEVELOPING CNS AND PROBABLY OTHER REGIONS OF THE EMBRYO.
CC -1- TISSUE SPECIFICITY: PROLIFERATING NEUROEPITHELIUM.
CC -1- DEVELOPMENTAL STAGE: CNS DEVELOPMENT.
CC -1- SIMILARITY: CONTAINS 34 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 5 ANK REPEATS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as their content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sdb.ch/announce/>
CC or send an email to license@sdb-sdb.ch).
CC -----
CC EMBL: X74760; CAA52776.1; -
DR HSSP: P00740; IIXA.
DR MGD: MGI:99460; Notch3.
DR InterPro: IPR002110; ANK.
DR InterPro: IPR000152; Asx_hydroxyl.
DR InterPro: IPR000561; EGF-like.

DR InterPro: IPR000742; EGF_2.
DR InterPro: IPR001881; EGF_Ca.
DR InterPro: IPR001438; EGF_II.
DR InterPro: IPR000800; Notch.
DR Pfam: PF000023; ank; 6.
DR Pfam: PF00006; EGF; 34.
DR Pfam: PF00066; notch; 3.
DR PRINTS: PR00010; EGFBLD.
DR SMART: SM00128; ANK; 5.
DR SMART: SM00179; EGF_Ca; 19.
DR SMART: SM00001; EGF_Like; 15.
DR PROSITE: PS50088; ANK_REPEAT; 4.
DR PROSITE: PS50297; ANK_REPEAT_REGION; 1.
DR PROSITE: PS01010; ASX_HYDROXYL; 18.
DR PROSITE: PS00022; EGF_1; 33.
DR PROSITE: PS01186; EGF_2; 27.
DR PROSITE: PS01187; EGF_Ca; 17.
KW Differentiation; Neurogenesis; Repeat; EGF-like domain; Transmembrane;
KW ANK repeat; Glycoprotein.
FT 1 1643
FT TRANSMEM 1644 1664
FT DOMAIN 1665 2318
FT DOMAIN 2242 2261
FT DOMAIN 39 78
FT DOMAIN 79 119
FT DOMAIN 120 157
FT DOMAIN 159 196
FT DOMAIN 198 235
FT DOMAIN 237 273
FT DOMAIN 275 313
FT DOMAIN 315 351
FT DOMAIN 352 390
FT DOMAIN 392 430
FT DOMAIN 432 468
FT DOMAIN 470 504
FT DOMAIN 508 544
FT DOMAIN 546 581
FT DOMAIN 583 619
FT DOMAIN 621 656
FT DOMAIN 658 694
FT DOMAIN 696 731
FT DOMAIN 735 771
FT DOMAIN 772 809
FT DOMAIN 811 848
FT DOMAIN 850 886
FT DOMAIN 888 923
FT DOMAIN 925 961
FT DOMAIN 963 999
FT DOMAIN 1001 1035
FT DOMAIN 1037 1083
FT DOMAIN 1085 1121
FT DOMAIN 1123 1159
FT DOMAIN 1161 1204
FT DOMAIN 1206 1245
FT DOMAIN 1247 1288
FT DOMAIN 1290 1326
FT DOMAIN 1336 1374
FT REPEAT 1388 1428
FT REPEAT 1429 1467
FT REPEAT 1468 1503
FT REPEAT 1839 1868
FT REPEAT 1872 1902
FT REPEAT 1906 1935
FT REPEAT 1939 1968
FT REPEAT 1972 2001
FT DISULFID 43 55
FT DISULFID 49 66
FT DISULFID 68 77
FT DISULFID 83 94
FT DISULFID 88 107
FT DISULFID 109 118
FT DISULFID 124 135

FT DISULFD 129 145 BY SIMILARITY.
 FT DISULFD 147 156 BY SIMILARITY.
 FT DISULFD 163 175 BY SIMILARITY.
 FT DISULFD 169 184 BY SIMILARITY.
 FT DISULFD 186 195 BY SIMILARITY.
 FT DISULFD 202 213 BY SIMILARITY.
 FT DISULFD 207 223 BY SIMILARITY.
 FT DISULFD 225 234 BY SIMILARITY.
 FT DISULFD 241 252 BY SIMILARITY.
 FT DISULFD 246 261 BY SIMILARITY.
 FT DISULFD 263 272 BY SIMILARITY.
 FT DISULFD 279 292 BY SIMILARITY.
 FT DISULFD 286 301 BY SIMILARITY.
 FT DISULFD 303 312 BY SIMILARITY.
 FT DISULFD 319 330 BY SIMILARITY.
 FT DISULFD 324 339 BY SIMILARITY.
 FT DISULFD 341 350 BY SIMILARITY.
 FT DISULFD 356 367 BY SIMILARITY.
 FT DISULFD 361 378 BY SIMILARITY.
 FT DISULFD 380 389 BY SIMILARITY.
 FT DISULFD 396 409 BY SIMILARITY.
 FT DISULFD 403 418 BY SIMILARITY.
 FT DISULFD 420 429 BY SIMILARITY.
 FT DISULFD 436 447 BY SIMILARITY.
 FT DISULFD 441 456 BY SIMILARITY.
 FT DISULFD 458 467 BY SIMILARITY.
 FT DISULFD 474 485 BY SIMILARITY.
 FT DISULFD 479 494 BY SIMILARITY.
 FT DISULFD 496 505 BY SIMILARITY.
 FT DISULFD 512 523 BY SIMILARITY.
 FT DISULFD 517 532 BY SIMILARITY.
 FT DISULFD 534 543 BY SIMILARITY.
 FT DISULFD 550 560 BY SIMILARITY.
 FT DISULFD 555 569 BY SIMILARITY.
 FT DISULFD 571 580 BY SIMILARITY.
 FT DISULFD 587 598 BY SIMILARITY.
 FT DISULFD 592 607 BY SIMILARITY.
 FT DISULFD 609 618 BY SIMILARITY.
 FT DISULFD 625 635 BY SIMILARITY.
 FT DISULFD 630 644 BY SIMILARITY.
 FT DISULFD 646 655 BY SIMILARITY.
 FT DISULFD 662 673 BY SIMILARITY.
 FT DISULFD 667 682 BY SIMILARITY.
 FT DISULFD 684 693 BY SIMILARITY.
 FT DISULFD 700 710 BY SIMILARITY.
 FT DISULFD 705 719 BY SIMILARITY.
 FT DISULFD 721 730 BY SIMILARITY.
 FT DISULFD 739 750 BY SIMILARITY.
 FT DISULFD 744 759 BY SIMILARITY.
 FT DISULFD 761 770 BY SIMILARITY.
 FT DISULFD 776 787 BY SIMILARITY.
 FT DISULFD 781 797 BY SIMILARITY.
 FT DISULFD 799 808 BY SIMILARITY.
 FT DISULFD 815 827 BY SIMILARITY.
 FT DISULFD 821 836 BY SIMILARITY.
 FT DISULFD 838 847 BY SIMILARITY.
 FT DISULFD 854 865 BY SIMILARITY.
 FT DISULFD 859 874 BY SIMILARITY.
 FT DISULFD 876 885 BY SIMILARITY.
 FT DISULFD 892 902 BY SIMILARITY.
 FT DISULFD 897 911 BY SIMILARITY.
 FT DISULFD 913 922 BY SIMILARITY.
 FT DISULFD 929 940 BY SIMILARITY.
 FT DISULFD 934 949 BY SIMILARITY.
 FT DISULFD 951 960 BY SIMILARITY.
 FT DISULFD 967 978 BY SIMILARITY.
 FT DISULFD 972 987 BY SIMILARITY.
 FT DISULFD 989 998 BY SIMILARITY.
 FT DISULFD 1005 1016 BY SIMILARITY.
 FT DISULFD 1010 1023 BY SIMILARITY.
 FT DISULFD 1025 1034 BY SIMILARITY.
 FT DISULFD 1041 1062 BY SIMILARITY.
 FT DISULFD 1056 1071 BY SIMILARITY.

FT DISULFD 1073 1082 BY SIMILARITY.
 FT DISULFD 1089 1100 BY SIMILARITY.
 FT DISULFD 1094 1109 BY SIMILARITY.
 FT DISULFD 1111 1120 BY SIMILARITY.
 FT DISULFD 1127 1138 BY SIMILARITY.
 FT DISULFD 1132 1147 BY SIMILARITY.
 FT DISULFD 1149 1158 BY SIMILARITY.
 FT DISULFD 1165 1183 BY SIMILARITY.
 FT DISULFD 1177 1192 BY SIMILARITY.
 FT DISULFD 1194 1203 BY SIMILARITY.
 FT DISULFD 1210 1223 BY SIMILARITY.

Query Match 16.7%; Score 70; DB 1; Length 2318;
 Best Local Similarity 37.7%; Pred. No. 38;
 Matches 23; Conservative 2; Mismatches 24; Indels 12; Gaps 1.

QY 6 LPRPAVPPPLMOPGPAHPVLSFLRPSMDLSARYSLPAPLSTSPISVSGRPD 65
 Db 2162 LNPVAVPLDMARLPPAPGPSFL-----LPLAPQOLLNPGAVSPQSRPP 2209
 QY 66 P 66
 Db 2210 P 2210

RESULT 15
 TEFS_CHICK STANDARD: PRF: 433 AA.
 ID 090701: 090702;
 AC 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DR 30-MAY-2000 (Rel. 39, Last annotation update)
 DE TRANSCRIPTIONAL ENHANCER FACTOR TEF-5 (CARDIAC-ENRICHED TEA DOMAIN
 TEFS OR DTEF1.
 GN Gallus gallus (Chicken).
 OS Gallus gallus (Chicken).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
 OC Gallus.
 OX NCBI_TaxID=9031;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Heart;
 RX MEDLINE=96215227; PubMed=8626520;
 RA Azakie A., Larkin S.B., Farrance I.K., Greeningloh G., Ordahl C.P.;
 RT "DTEF-1, a novel member of the transcription enhancer factor-1
 (TEF-1) multigene family.";
 RL J. Biol. Chem. 271:8260-8265(1996).
 CC -1- FUNCTION: SEQUENCE SPECIFIC M-CAT-BINDING FACTOR.
 CC -1- SUBCELLULAR LOCATION: NUCLEAR (POTENTIAL).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; DTEF-1A (SHOWN HERE) AND DTEF-
 1B: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: HIGH LEVELS IN CARDIAC MUSCLE, LOW IN SKELETAL
 MUSCLE. INTERMEDIATE LEVELS IN GIZZARD AND LUNG, LOW LEVELS IN
 KIDNEY.
 CC -1- SIMILARITY: CONTAINS A 'TEA' DNA-BINDING DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: 046127; AAC59786.1; ALT_INIT.
 DR EMBL: 046128; AAC59787.1; ALT_INIT.
 DR InterPro: IPR000818; TEA.
 DR Pfam: PF01285; TEA; 1.
 DR PRINTS: PR00065; TEADOMAIN.
 DR SMART: SM00426; TEA; 1.
 DR PROSITE: PS00534; TEA_DOMAIN; 1.
 KW Transcription regulation; Activator; DNA-binding; Nuclear protein;

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:43 ; Search time 40.45 Seconds

(without alignments) 285.674 Million cell updates/sec

Title: US-09-234-208b-1

Perfect score: 418

Sequence: 1 GTHSLPPRAVAVPPLRMQP.....VGRGPPDAHAVVNLRYEG 79

Scoring table: BLOSUM62

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_17.*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	406	97.1	419	4	Q9UK79
2	85	20.3	327	11	Q99JK6
3	84.5	20.2	434	5	Q9GJ05
4	84	20.1	995	11	Q35615
5	81.5	19.5	816	11	Q70474
6	81	19.4	1006	10	Q82761
7	81	19.4	1006	10	Q82761
8	78.5	18.8	955	12	Q9IMW0
9	78	18.7	122	12	Q9YLRO
10	78	18.7	763	2	Q9XDH2
11	78	18.7	2319	11	Q9R172
12	77.5	18.5	217	3	Q9P3L0
13	77.5	18.5	487	5	Q9V197
14	77.5	18.5	1212	2	Q9LIC8
15	77	18.4	122	12	Q36612
16	77	18.4	503	5	Q17585
17	76.5	18.3	1110	5	Q09493
18	76	18.2	309	3	Q9P6R1
19	76	18.2	356	10	Q9SAK1

20	76	18.2	1044	4	Q9NXC6	Q9nxc6 homo sapien
21	75	17.9	189	12	Q08404	Q08404 soybean dwa
22	75	17.9	189	12	Q87034	Q87034 soybean dwa
23	75	17.9	295	2	Q9KX07	Q9KX07 streptomyc
24	75	17.9	1081	4	Q9HBF3	Q9HBF3 homo sapien
25	75	17.9	1952	3	Q9P6T1	Q9P6T1 neurospora
26	74.5	17.8	894	10	Q9SUS3	Q9SUS3 arabidopsis
27	74.5	17.8	894	10	Q9M6S0	Q9M6S0 arabidopsis
28	74	17.7	384	11	Q9D789	Q9D789 mus musculu
29	73.5	17.6	574	4	Q9BT08	Q9BT08 homo sapien
30	73.5	17.6	1822	4	Q9P1V7	Q9P1V7 homo sapien
31	73.5	17.6	2321	12	Q9DGT6	Q9DGT6 turkey harp
32	73	17.5	123	12	Q9WLK2	Q9WLK2 hepatitis e
33	73	17.5	639	4	Q9H9M1	Q9H9M1 homo sapien
34	73	17.5	1180	5	Q9VRM2	Q9VRM2 drosophila
35	73	17.5	1520	4	Q9Y4E0	Q9Y4E0 homo sapien
36	73	17.5	1664	4	Q9BZES	Q9BZES homo sapien
37	73	17.5	3503	5	Q24292	Q24292 drosophila
38	72.5	17.3	518	2	Q53738	Q53738 mycobacteri
39	72.5	17.3	715	2	Q9F9V7	Q9F9V7 mycobacteri
40	72.5	17.3	1410	5	Q9GRM9	Q9GRM9 leishmania
41	72.5	17.3	1709	4	Q15047	Q15047 homo sapien
42	72	17.2	122	12	Q71146	Q71146 hepatitis e
43	72	17.2	489	4	Q9B0C3	Q9B0C3 homo sapien
44	72	17.2	680	12	Q90CX4	Q90CX4 chayote mos
45	72	17.2	1914	13	Q91008	Q91008 gallus gall

ALIGNMENTS

RESULT	1	PRELIMINARY:	PRT:	419 AA.
Q9UK79	Q9UK79	Q9UK79		
ID	Q9UK79	Q9UK79		
AC	Q9UK79	Q9UK79		
DT	01-MAY-2000 (TREMBLrel. 13, Created)			
DR	01-MAR-2001 (TREMBLrel. 16, Last sequence update)			
DI	01-JUN-2001 (TREMBLrel. 17, Last annotation update)			
DE	HERSTATIN.			
GN	HER-2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=99415951; PubMed=10485918;			
RA	Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;			
RT	"The HER-2/neu receptor tyrosine kinase gene encodes a secreted			
RT	autoinhibitor.";			
RL	Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RA	Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Hennen W.D.;			
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.			
DR	EMBL: AF177761; AAD56009.2; -			
DR	InterPro: IPR000494; EGFR_L.			
DR	InterPro: IPR002174; Furin-Like.			
DR	Pfam: PF00757; Furin-Like; 1.			
DR	Pfam: PF01030; Recep_L_domain; 1.			
DR	SMART: SMO0261; FU; 1.			
DR	SEQUENCE 419 AA; 45472 MW; FEClBE347E2D030C CMC64;			
QY	Query Match	97.1%; Score 406; DB 4; Length 419;		
DB	Best Local Similarity	97.5%; Pred. No. 4.2e-34;		
QY	Matches 77; Conservative 1; Mismatches 1; Indels 0; Gaps 0;			
DB	1 GTHSLPPRAVAVPPLRMQPGPAHVLSTLRPSMDVSAFYSLPLAPSPISPVSV 60			
QY				
DB	341 GTHSLPPRAVAVPPLRMQPGPAHVLSTLRPSMDVSAFYSLPLAPSPISPVSV 400			
QY				
DB	61 GRCPPDAHAVVNLRYEG 79			

Db 401 GRGPDPAHVAVDLSRYEG 419

RESULT 2
ID 099JK6 PRELIMINARY; PRT; 327 AA.
AC 099JK6;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE UNKNOWN (PROTEIN FOR IMAGE:3591061) (FRAGMENT).
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (APR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC006054; AAH06054.1; -
FT NON_TER 1
SQ SEQUENCE 327 AA; 33661 MW; 27917F16D583E774 CRC64;

Query Match 20.3%; Score 85; DB 11; Length 327;
Best Local Similarity 42.4%; Pred. No. 0.35;
Matches 28; Conservative 4; Mismatches 28; Indels 6; Gaps 4;

OY 1 GTHSLPRPAVPLRMQGPAPHVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPSV 60
DB 88 GASGSCSP-PGRCPAPLRPDSPPTNPAMSPRRPARGLDAA--SSP--PLEGSPSPSP-PA 141
OY 61 GRGPD 66
DB 142 GLSP 147

RESULT 3
ID 09GUL5 PRELIMINARY; PRT; 434 AA.
AC 09GUL5;
DT 01-MAR-2001 (TREMBLrel. 16, Created)
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 49.7 KDA PROTEIN.
OS Theileria parva.
OC Eukaryota; Alveolata; Apicomplexa; Piroplasmida; Theileriidae;
OC Theileria.
OX NCBI_TaxID=5875;
RN [1]
RP SEQUENCE FROM N.A.
RA Bishop R., Godlight E., Nene V., Morzaria S., Musoke A., Sohanpal B.;
RT "Polymorphic open reading frames encoding secretory proteins are
located less than 3 kilobases from Theileria parva telomeres."
RL Mol. Biochem. Parasitol. 110:359-371(2000).
DR EMBL: AF225701; AAC28022.1; -
DR InterPro: IPR002965; P_rich_extensn.
DR PRINTS: PR01217; PRICHEXTENSN.
KW Hypothetical protein.
SQ SEQUENCE 434 AA; 49670 MW; 0FA6586F4021A22F CRC64;

Query Match 20.2%; Score 84.5; DB 5; Length 434;
Best Local Similarity 31.0%; Pred. No. 0.53;
Matches 26; Conservative 11; Mismatches 34; Indels 13; Gaps 4;

OY 4 SLPRPAVPLV--PLRMQGP-PAHPVLSFLRPSMDLVSAF-----SLPLAPLSPTSPV 55
DB 159 ALYRPPIOPIDPLPDGPHAPRP-----PRYPVSGTSPYHPYAPSPVPQIPPPS 213
OY 56 SPVSVGRGPDPAHVAVDLSRYEG 79

Db 214 THVVPETQPOPOQVPOQVPOQYQYQY 237

RESULT 4
ID 035615 PRELIMINARY; PRT; 995 AA.
AC 035615;
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE FRIEND OF GATA-1 (FOG).
GN ZFPML OR FOG.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97373824; PubMed=9230307;
RA Tsang A.P., Visvader J.E., Turner C.A., Fujiwara Y., Yu C.,
RA Weiss M.J., Crossley M., Orkin S.H.;
RT "FOG, a multitype zinc finger protein, acts as a cofactor for
transcription factor GATA-1 in erythroid and megakaryocytic
differentiation."
RT Cell 90:109-119(1997).
RL -1- SUBCELLULAR LOCATION: NUCLEAR (BY SIMILARITY).
CC EMBL: AF006492; AAC53292.1; -
DR MGD: I095400; Zfpml.
DR InterPro: IPR001005; MYB_DNA_bind.
DR InterPro: IPR000822; Znf-C2H2.
DR Pfam: PF00096; zfc-C2H2; 9.
DR PRINTS: PR00048; ZINCINGER.
DR SMART: SM00355; Znf.C2H2; 5.
DR PROSITE: PS00037; MYB_1; UNKNOWN_1.
DR PROSITE: PS00028; ZINC_FINGER_C2H2_1; UNKNOWN_2.
DR PROSITE: PS00157; ZINC_FINGER_C2H2_2; 2.
KW DNA-binding; Nuclear protein; Zinc-finger.
SQ SEQUENCE 995 AA; 105983 MW; 293255B28151ECB8 CRC64;

Query Match 20.1%; Score 84; DB 11; Length 995;
Best Local Similarity 33.3%; Pred. No. 1.4;
Matches 23; Conservative 10; Mismatches 28; Indels 8; Gaps 2;

OY 7 PRPAVPLVPLRMQGPAPHVLSFLRPSMDLVSAFYSLPLAPLSPTSPVSPVGRGP-- 64
DB 766 PPPAGPAPVPPVPSPTALFSPSPR-----GSASGAPAPALSPSPVDPGPIIDLSKRPR 820
OY 65 -DPDAHVA 72
DB 821 QSPDAPFAL 829

RESULT 5
ID 070474 PRELIMINARY; PRT; 816 AA.
AC 070474;
DT 01-AUG-1998 (TREMBLrel. 07, Created)
DT 01-AUG-1998 (TREMBLrel. 07, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE NEUROCAN (FRAGMENT).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SPRAGUE-DAWLEY; TISSUE=HIPPOCAMPUS;
RA Zachmann-Brand B., Schaller H.C.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF060879; AAC15766.1; -
DR HSSP: P00740; 1EDM.

DR InterPro: IPR000152; Asx_hydroxyl.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR000742; EGF_2.
 DR InterPro: IPR001881; EGF_Ca.
 DR InterPro: IPR001438; EGF_II.
 DR InterPro: IPR001304; Lectin_C.
 DR InterPro: IPR000538; Link.
 DR Pfam: PF00008; EGF_2.
 DR Pfam: PF00193; Xlink_1.
 DR PRINTS: PRO0010; EGFLOOD.
 DR PRINTS: PRO1265; LINKMODULE.
 DR ProDom: PD000918; Link_1.
 DR SMART: SM00179; EGF_CA; 1.
 DR SMART: SM00045; EGF_like; 1.
 DR PROSITE: PS00010; ASX_HYDROXYL; UNKNOWN_1.
 DR PROSITE: PS50041; C_TYPE_LECTIN_2; 1.
 DR PROSITE: PS00022; EGF_1; UNKNOWN_3.
 DR PROSITE: PS01186; EGF_2; 1.
 DR PROSITE: PS01187; EGF_CA; 1.
 DR PROSITE: PS01241; LINK; 1.
 DR Calcium binding; EGF-like domain; Glycoprotein; Repeat.
 FT NON_TER 1
 FT NON_TER 816
 SQ SEQUENCE 816 AA; 85578 MW; 957F5917AD10616E CRC64;

Query Match 19.5%; Score 81.5; DB 11; Length 816;
 Best Local Similarity 35.0%; Pred. No. 2.1;
 Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

OY 4 SLPPRAAVPVPLRMQPG---PAHPVLSFLR-----PSMDVSAFYSLPLAPLS--PT 51
 DB 342 SSTPSALSAVSLQSPGSGSPDFIVAMLRAPKMLPLPSTLVVSPPLSPASPPLPS 401
 OY 52 SVP---ISPSVSGRPPDP 67
 DB 402 SVPEQAVRPVSFG-AEDPE 420

RESULT 6
 082761
 AC 082761 PRELIMINARY; PRT; 200 AA.
 DT 01-NOV-1998 (TREMBLrel. 08, Created)
 DT 01-NOV-1998 (TREMBLrel. 08, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE PUTATIVE BLUE COPPER-BINDING PROTEIN.
 GN T16B12.14.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Rounsley S.D., Lin X., Kaul S., Shea T.P., Fujii C.Y., Mason T.M.,
 RA Shen M., Rounsley C.M., Fraser C.M., Somerville C.R., Venter J.C.,
 RT "Arabidopsis thaliana chromosome II BAC T16B12 genomic sequence."
 RL Submitted (OCT-1998) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC005311; AAC63847.1; -
 DR HSSP: P00303; 2CBP.
 DR InterPro: IPR003245; Cu_bind_like.
 DR Pfam: PF02298; Cu_bind_like; 1.
 DR ProDom: PD003122; Cu_bind_like; 1.
 SQ SEQUENCE 200 AA; 21475 MW; E669011C997E349C CRC64;

Query Match 19.4%; Score 81; DB 10; Length 200;
 Best Local Similarity 34.3%; Pred. No. 0.55;
 Matches 23; Conservative 5; Mismatches 23; Indels 16; Gaps 2;

OY 9 PAAPVPLRMQGPAPHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPSVSGRPPDA 68
 DB 132 PYAAPV-----PGVAPRPSFSFSSPS-----QSPLASPVNHAIVQRMGSPAP 175
 OY 69 HVAANLS 75
 DB 176 HSAASNS 182

RESULT 7
 091MQ1
 ID 091MQ1 PRELIMINARY; PRT; 1006 AA.
 AC 091MQ1
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE F7H2.17 PROTEIN.
 GN F7H2.17.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 OC eustosids II; Brassicales; Brassicaceae; Arabidopsids.
 OX NCBI_TaxID=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. COLUMBIA;
 RA Liu S.X., Sakano H., Yu G., Etgu P., Lee J., Lenz C., Pham P.,
 RA Toriumi M., Chan A., Chung M., Goldsmith A., Liu A., Smith A.,
 RA Vaysberg M., Altafi H., Brooks S., Buehler E., Chao Q., Conn L.,
 RA Conway A.B., Hansen N.F., Johnson-Hopson C., Khan S., Kim C., Lam B.,
 RA Miranda M., Nguyen M., Palm C.J., Shinn P., Southwick A., Davis R.W.,
 RA Ecker J.R., Federspiel N.A., Theologis A.,
 RT "The sequence of BAC F7H2 from Arabidopsis thaliana chromosome 1."
 RL Submitted (JUL-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AC034256; AAF82153.1; -
 DR InterPro: IPR002965; P-rich_extensions.
 DR PRINTS: PRO1217; PRICHEXTENS.
 SQ SEQUENCE 1006 AA; 103943 MW; C9FB49F9930C238D CRC64;

Query Match 19.4%; Score 81; DB 10; Length 1006;
 Best Local Similarity 31.4%; Pred. No. 2.9;
 Matches 22; Conservative 5; Mismatches 21; Indels 22; Gaps 2;

OY 6 LPPRAAVPVPLRMQGPAPHPVLSFLRPSMDVSAFYSLPLAPLSPTSPISPSVSGRPPD 65
 DB 336 LPPPATLPPPLPLPPPP-----SLPVPFCSP---PPPIIYNGAPP 373
 OY 66 PDAHVAANLS 75
 DB 374 PPCVTICVOVS 383

RESULT 8
 091MY0
 ID 091MY0 PRELIMINARY; PRT; 955 AA.
 AC 091MY0
 DT 01-OCT-2000 (TREMBLrel. 15, Created)
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
 DE LATENT NUCLEAR ANTIGEN EBNA-3A.
 GN EBNA-3A.
 OS Cerepithicline herpesvirus 15.
 OC Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
 OC Gammaherpesvirinae; Lymphocryptovirus.
 OX NCBI_TaxID=104228;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LC18664;
 RX MEDLINE=20304984; PubMed=10846073;
 RA Jiang H., Cho Y.-G., Wang F.,
 RT "Structural, Functional, and Genetic Comparisons of Epstein-Barr Virus
 Nuclear Antigen 3A, 3B, and 3C Homologs Encoded by the Rhesus

Query Match	Best Local Similarity	Score	DB	Length	2319;
SMART: SM00179; EGF_CA; 20.					
DR SMART: SM00001; EGF-like; 12.					
DR SMART: SM00004; NL; 3.					
DR PROSITE; PS50088; ANK_REPEAT; 4.					
DR PROSITE; PS50297; ANK_REPEAT_REGION; 1.					
DR PROSITE; PS00010; ASX_HYDROXYL; 1.					
KW EGF-like domain; Glycoprotein; Hydroxylation.					
SO SEQUENCE 2319 AA; 244298 MW; 2438CA02D/C3283D CRC64;					
Query Match	18.7%;	Score 78;	DB 11;	Length 2319;	
Best Local Similarity	39.3%;	Pred. No. 14;			
Matches 24;	Conservative 2;	Mismatches 23;	Indels 12;	Gaps 1;	
OY 6	LPRPAVAVPPLRMQGPAPHVLSFLRPSMDLVSATYSLLPLAPLSPTSPVSISSVSGRGPD 65				
DB 2163	LINPAVAVPLDMARLPAPPAPGPSFL-----LPLAGSOLLNPATVPSPHERPP 2210				
OY 66	P 66				
DB 2211	P 2211				
RESULT 12					
O9P3L0	PRELIMINARY;	PRT;	217	AA.	
AC O9P3L0;					
DT 01-OCT-2000 (TREMBLrel. 15, Created)					
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)					
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)					
DE RELATED TO FINGER PROTEIN XFG 68.					
GN B15120.10.					
OS Neurospora crassa.					
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;					
OC Sordariales; Sordariaceae; Neurospora.					
OX NCBI_TaxID=5141;					
RN [1]					
RP SEQUENCE FROM N.A.					
RA Schulte U., Aign V., Hohnsels J., Brandt P., Fartmann B., Holland R.,					
RA Nyakatura G., Mewes H.W., Mannhaupt G.;					
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.					
RN [2]					
RP SEQUENCE FROM N.A.					
RA German Neurospora genome project;					
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.					
DR EMBL; AL389900; CAB97455.1;					
SO SEQUENCE 217 AA; 23786 MW; 100F7501806C0AC0 CRC64;					
Query Match	18.5%;	Score 77.5;	DB 3;	Length 217;	
Best Local Similarity	37.5%;	Pred. No. 1.4;			
Matches 21;	Conservative 5;	Mismatches 21;	Indels 9;	Gaps 2;	
OY 7	PRPAVAVPPLRMQGPAPHVLSFLRPSMDLVSATYSLLPLAPLSPTSPVSISSVSGR 62				
DB 109	PRPSPASIPRRTCPPIPRP-----PS---TSALVLLVSPSPSSSPILPCSPRR 155				
RESULT 13					
O9VI97	PRELIMINARY;	PRT;	487	AA.	
AC O9VI97;					
DT 01-MAY-2000 (TREMBLrel. 13, Created)					
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)					
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)					
DE CGI4604 PROTEIN.					
GN CGI4604.					
OS Drosophila melanogaster (Fruit fly).					
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;					
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;					
OC Ephydroidea; Drosophilidae; Drosophila.					
OX NCBI_TaxID=7227;					
RN [1]					

RP SEQUENCE FROM N.A.
RC STRAIN-BERKELEY;
RC MEDLINE-20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.-H.C., Blaise R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abiri J.F., Agbayan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Hoch A., Dunkov B.C., Dunn P.,
RA Dubin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Fouts C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Honck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Iqbalwan C.,
RA Jaitai M., Kalusz F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lal Z.,
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Paclab J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wassarman D.A., Weisbach J., Weissbach J.,
RA Williams S.M., Woodage T., Wolley K.C., Wu D., Yang S., Yao Q.A.,
RA Ye J., Yen R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RA "The genome sequence of *Drosophila melanogaster*.";
RL Science 287:2185-2195(2000).
DR EMBL; AE003672; AAF54028.1; -
DR FlyBase; FBgn0037499; CG14604.
SQ SEQUENCE 487 AA; 52727 MW; E0CBDE4612FB6B4A CRC64;

Query Match 18.5%; Score 77.5; DB 5; Length 487;
Best Local Similarity 30.9%; Pred. No. 3.1;
Matches 21; Conservative 9; Mismatches 23; Indels 15; Gaps 2;

QY 7 PRPAVP---VPLRMQGPAPHPVLS-----FLRPSWDLVSFAFSLPLAPLSPT 51
DB 121 PAPAFAPIAAGVPMVPGQSPVFPFAVSPAAPAAVAPVAPAPAAVAPVAPVAPAPAPS 180
QY 52 SVRISPV 59
DB 181 VVPVAPVA 188

RESULT 14
Q9LIC8 PRELIMINARY; PRI; 1212 AA.
AC Q9LIC8: 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, last annotation update)
DE PUTATIVE NITROREDUCTASE.
GN SC11.10C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinomycetia;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomycetes.
NX NCBI_Taxid=1902;
NN [1]

Db 112 v 112

Search completed: April 11, 2002, 09:35:06
Job time: 323 sec

RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Oliver K., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RA Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RX MEDLINE-97000351; PubMed-8843436;
RA Redenbach M., Kieser H.M., Denaplatte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL157953; CAB/6073.1; -;
DR InterPro; IPR003200; DBI_PRT.
DR InterPro; IPR000415; Nitroreductase.
DR InterPro; IPR002965; P_Rich_extensn.
DR Pfam; PF02277; DBI_PRT; 1.
DR Pfam; PF00881; Nitroreductase; 1.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 1212 AA; 124280 MW; E898B903F78D235A CRC64;

Query Match 18.5%; Score 77.5; DB 2; Length 1212;
Best Local Similarity 36.7%; Pred. No. 8;
Matches 22; Conservative 7; Mismatches 28; Indels 3; Gaps 2;

OY 9 PAAVPPLEMOGPAPHPVSLRPSWDLVSATFSLPLAPLSPTSVISPSVGRGPPDA 68
DB 488 PAPVPAPVAPQEPHPAPQVGFPLVDGGVPT--THLAPTPPEAVLVPEETAAE-PEVYA 544

RESULT 15
036612 PRELIMINARY; PRT; 122 AA.
ID 036612;
AC 036612;
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-JUN-2001 (TRENBLREL. 17, Last annotation update)
DE HYPOTHETICAL 12.5 KDA PROTEIN (ORF-3).
OS Swine hepatitis E virus.
OC Viruses; ssRNA positive-strand viruses, no DNA stage.
OX NCBI_TaxID=63421;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MENG;
RX MEDLINE-97420774; PubMed-9275216;
RA Meng X.J., Purcell R.H., Halbur P.G., Lehman J.R., Webb D.M.,
RA Tsareva T.S., Haynes J.S., Thacker B.J., Emerson S.U.;
RT "A novel virus in swine is closely related to the human hepatitis E
virus";
RT Proc. Natl. Acad. Sci. U.S.A. 94:9860-9865(1997).
RL EMBL; AF082843; AAC97209.1; -;
DR InterPro; IPR003384; HEV_ORF2.
DR Pfam; PF02444; HEV_ORF2; 1.
SQ SEQUENCE 122 AA; 12466 MW; 77602F9048E7B12A CRC64;

Query Match 18.4%; Score 77; DB 12; Length 122;
Best Local Similarity 36.1%; Pred. No. 0.85;
Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 3;

OY 1 GTHTSLPRAAVPVPLRMQGPAPHPVSLRPSWDLVSATFSLPLAPL--SPTSVISPS 57
DB 55 GTTGLILSPS--PSPFIQPTPSLP-MSFHNPGLEFALDSRAPLAPLGVTSPSAPPLPP 111
OY 58 v 58

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:28:23 ; Search time 20.27 Seconds
(without alignments)
87.704 Million cell updates/sec

Title: US-09-234-208B-1

Perfect score: 418

Sequence: 1 GTHSLPRLPAAPVPLRMQP.....VGRGPDPAHVAVLSRYEG 79

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PCITUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfills1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	81.5	19.5	1257	1 US-08-340-428B-49	Sequence 49, Appl
2	72	17.2	440	2 US-08-430-286A-9	Sequence 9, Appl
3	71.5	17.1	604	2 US-08-468-576B-12	Sequence 12, Appl
4	71.5	17.1	604	2 US-08-468-579B-12	Sequence 12, Appl
5	71.5	17.1	604	3 US-08-468-577B-12	Sequence 12, Appl
6	71.5	17.1	2441	1 US-08-194-468-2	Sequence 2, Appl
7	71.5	17.1	2441	3 US-08-961-739-2	Sequence 2, Appl
8	69.5	16.6	432	1 US-08-615-170-21	Sequence 21, Appl
9	69.5	16.6	433	1 US-08-615-170-19	Sequence 21, Appl
10	68.5	16.4	115	4 US-09-461-697-58	Sequence 58, Appl
11	68.5	16.4	132	4 US-09-461-697-52	Sequence 54, Appl
12	68.5	16.4	139	4 US-09-461-697-52	Sequence 52, Appl
13	68.5	16.4	139	4 US-09-461-697-48	Sequence 48, Appl
14	67.5	16.1	123	3 US-08-840-316-3	Sequence 9, Appl
15	67.5	16.1	123	3 US-08-478-507-9	Sequence 9, Appl
16	67.5	16.1	123	4 US-08-809-523-3	Sequence 3, Appl
17	67.5	16.1	123	4 US-09-128-275A-9	Sequence 9, Appl
18	67.5	16.1	123	4 US-08-471-971-3	Sequence 3, Appl
19	67.5	16.1	123	5 PCT-US93-08849A-3	Sequence 3, Appl
20	67.5	16.1	123	5 PCT-US93-08849-3	Sequence 3, Appl
21	67.5	16.1	124	4 US-08-240-049B-19	Sequence 19, Appl
22	67.5	16.1	124	4 US-08-542-634-21	Sequence 21, Appl
23	67.5	16.1	124	4 US-08-477-293-21	Sequence 21, Appl
24	67.5	16.1	124	5 PCT-US95-13703-21	Sequence 21, Appl
25	67.5	16.0	174	4 US-08-818-112-143	Sequence 143, App
26	66.5	15.9	1487	4 US-08-760-489-2	Sequence 2, Appl
27	66.5	15.9	1487	2 US-08-760-489-4	Sequence 4, Appl

28	66	15.8	325	4 US-09-041-886-33	Sequence 33, Appl
29	66	15.8	754	2 US-08-525-864A-2	Sequence 2, Appl
30	66	15.8	1447	4 US-09-041-886-25	Sequence 25, Appl
31	66	15.8	1447	5 PCT-US94-05277-2	Sequence 2, Appl
32	65.5	15.7	1257	4 US-09-220-641-3	Sequence 3, Appl
33	65	15.6	267	4 US-08-818-112-142	Sequence 142, App
34	65	15.6	1291	4 US-09-150-460B-10	Sequence 10, Appl
35	65	15.6	1291	4 US-09-220-641-5	Sequence 5, Appl
36	64.5	15.4	124	1 US-08-240-049B-20	Sequence 20, Appl
37	64.5	15.4	124	4 US-08-542-634-22	Sequence 22, Appl
38	64.5	15.4	124	4 US-08-477-292-22	Sequence 22, Appl
39	64.5	15.4	124	5 PCT-US95-13703-22	Sequence 22, Appl
40	64.5	15.4	543	2 US-08-469-412A-7	Sequence 7, Appl
41	64.5	15.4	543	2 US-09-021-715-7	Sequence 7, Appl
42	64.5	15.4	548	4 US-08-469-412A-2	Sequence 2, Appl
43	64.5	15.4	548	4 US-09-021-715-2	Sequence 2, Appl
44	64.5	15.4	1253	1 US-08-252-966B-12	Sequence 12, Appl
45	64.5	15.4	1261	1 US-08-252-966B-18	Sequence 18, Appl

ALIGNMENTS

RESULT 1
US-08-340-428B-49
; Sequence 49, Application US/08340428B
; Patent No. 5648465
; GENERAL INFORMATION:
; APPLICANT: MARGOLIS, Richard U.
; APPLICANT: RAUCH, Dwe
; APPLICANT: MARGOLIS, Renee K.
; TITLE OF INVENTION: CLONING, EXPRESSION AND USES FOR A
; TITLE OF INVENTION: NEUROCAN AS A CHONDROITIN SULFATE PROTEOGLYCAN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: Browdy and Neimark
; STREET: 419 Seventh Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20004
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/340,428B
; FILING DATE: 14 No. 5648465ember 1994
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/922,911
; FILING DATE: 03 August 1992
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Browdy, Roger L.
; REGISTRATION NUMBER: 25,618
; REFERENCE/DOCKET NUMBER: Margolis-1A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-628-5197
; TELEFAX: 202-737-3528
; INFORMATION FOR SEQ ID NO: 49:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1257 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-08-340-428B-49
Query Match 19.5%; Score 81.5; DB 1; Length 1257;
Best Local Similarity 35.0%; Pred. No. 0.3;

Matches 28; Conservative 10; Mismatches 25; Indels 17; Gaps 5;

QY 4 SLIPRAAVPVPLRMQPG---PAHPVLSLR-----PSMDLYSAFYSLPLAPLS--PT 51

Db 610 SSTSPSEALSAVSIQASPDGSPDPVAMLRAPKMLLPSTLVPNVSPILPSAPSLPS 669

QY 52 SVP---ISPVSVGRGPPD 67

Db 670 SVPEEQAVRPVSFG-AEDPE 688

RESULT 2

US-08-430-286A-9
; Sequence 9, Application US/08430286A
; Patent No. 6225080

GENERAL INFORMATION:

APPLICANT: UHL, George R.

APPLICANT: Eppler, C. Mark

APPLICANT: Wang, Jai-Bel

TITLE OF INVENTION: Mu-Subtype Opioid Receptor

NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Darby & Darby PC

STREET: 805 Third Avenue

CITY: New York

STATE: New York

COUNTRY: US

ZIP: 10022

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/430, 286A

FILING DATE: 28-APR-1995

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Robinson, Joseph R.

REGISTRATION NUMBER: 33,448

REFERENCE/DOCKET NUMBER: 0646/1A843-US5

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-527-7700

TELEFAX: 212-753-6237

TELEX: 236687

INFORMATION FOR SEQ ID NO: 9:

SEQUENCE CHARACTERISTICS:

LENGTH: 440 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

FRAGMENT TYPE: N-terminal

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: OPB-R

US-08-430-286A-9

Query Match 17.2%; Score 72; DB 4; Length 440;

Best Local Similarity 45.2%; Pred. No. 1;

Matches 19; Conservative 5; Mismatches 10; Indels 8; Gaps 3;

QY 2 THSLPRAVVPVPLRMQPGPAHPVLSLRPSMDLYSAFYSL 43

Db 32 TASPSAPSWTSPSP---RPGPAHP---FLQPPMAV---ALMSL 65

RESULT 3

US-08-468-576B-12
; Sequence 12, Application US/08468576B
; Patent No. 5953345

GENERAL INFORMATION:

APPLICANT: Rabin, Daniel

TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

ADDRESSEE: Sprung Kramer Schaefer & Briscoe

STREET: 660 White Plains Road

CITY: Tarrytown

STATE: New York

COUNTRY: USA

ZIP: 10591-5144

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage

COMPUTER: Apple Macintosh

OPERATING SYSTEM: System 7.5

SOFTWARE: Wordperfect

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468, 576B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/239,276

FILING DATE: 05-MAY-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/872,646

FILING DATE: 08-JUN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/715,181

FILING DATE: 14-JUN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/441,703

FILING DATE: 04-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/312,543

FILING DATE: 17-FEB-1989

ATTORNEY/AGENT INFORMATION:

NAME: Kurt G. Briscoe

REGISTRATION NUMBER: 33,141

REFERENCE/DOCKET NUMBER: MDI 251.7-KGB

TELECOMMUNICATION INFORMATION:

TELEPHONE: (914) 332-1700

TELEFAX: (914) 332-1844

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:

LENGTH: 604 amino acids

TYPE: amino acid

TOPOLOGY: linear

US-08-468-576B-12

Query Match 17.1%; Score 71.5; DB 2; Length 604;

Best Local Similarity 34.8%; Pred. No. 1.8;

Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 10 AAVPVPLRMQPGPAHPVLSLRPSMDLYSAFYSLPAPLSPTSVIPSV--SVGRGPPD 67

Db 421 AAQPLNLSRKPATAPKPTSPPTONLFPAKTSPTVLPKSSIP-SPIGSLGRGSLD 479

QY 68 AHVAVN 73

Db 480 ILSSLN 485

US-08-468-579B-12
; Sequence 12, Application US/08468579B
; Patent No. 5981700

GENERAL INFORMATION:

APPLICANT: Rabin, Daniel

TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS

TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING

NUMBER OF SEQUENCES: 19

CORRESPONDENCE ADDRESS:

```

ADDRESS: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,579B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.5-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-579B-12

Query Match          17.1%; Score 71.5; DB 2; Length 604;
Best Local Similarity 34.8%; Pred. No. 1.8;
Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 10 AAVPPLMOPGPAHVPVLSFLRPSMDLVSAFYSLPLAPLFTSPVISPV--SVGRGPPD 67
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 421 AAOPLNLSRKRTPKXSPSPNTONLFPASKTSPVNLPNKSSIP-SPIGSLGRGSSLD 479

QY 68 AHVAVN 73
    :|:|
Db 480 ILSSIN 485

RESULT 5
US-08-468-579B-12
; Sequence 12, Application US/08468577B
; Patent No. 6001804
; GENERAL INFORMATION:
; APPLICANT: Rabin, Daniel
; TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
; TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sprung Kramer Schaefer & Briscoe
; STREET: 660 White Plains Road
; CITY: Tarrytown
; STATE: New York
; COUNTRY: USA
```

```

ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 MB storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: WordPerfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,577B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kurt G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.8-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 604 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-579B-12

Query Match          17.1%; Score 71.5; DB 3; Length 604;
Best Local Similarity 34.8%; Pred. No. 1.8;
Matches 23; Conservative 13; Mismatches 27; Indels 3; Gaps 2;

QY 10 AAVPPLMOPGPAHVPVLSFLRPSMDLVSAFYSLPLAPLFTSPVISPV--SVGRGPPD 67
    |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db 421 AAOPLNLSRKRTPKXSPSPNTONLFPASKTSPVNLPNKSSIP-SPIGSLGRGSSLD 479

QY 68 AHVAVN 73
    :|:|
Db 480 ILSSIN 485

RESULT 6
US-08-194-468-2
; Sequence 2, Application US/08194468
; Patent No. 5750336
; GENERAL INFORMATION:
; APPLICANT: Montminy, Marc R.
; TITLE OF INVENTION: ASSAYS FOR THE IDENTIFICATION OF
; TITLE OF INVENTION: COMPOUNDS WHICH INHIBIT ACTIVATION OF CAMP AND MITOGEN
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pretty, Schroeder, Brueggemann & Clark
; STREET: 444 South Flower Street, Suite 2000
; CITY: Los Angeles
; STATE: California
; COUNTRY: USA
; ZIP: 90071
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
```

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/08/194,468
APPLICATION NUMBER: US/08/194,468
FILING DATE: 10-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Reiter, Stephen E.
REGISTRATION NUMBER: 31,192
REFERENCE/DOCKET NUMBER: P41 9672
TELECOMMUNICATION INFORMATION:
TELEPHONE: (619)-546-4737
TELEFAX: (619)-546-9392
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2441 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-194-468-2

Query Match 17.1%; Score 71.5; DB 1; Length 2441;
Best Local Similarity 36.5%; Pred. No. 10;
Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

OY 4 SLPPRAAVPVLRMQPGAHVLSFLRPSMDLVSAFSLPLAPLSPTSPVSVGNG 63
DB 843 SOLPCPPTVQSLPHPTPPASTAGM--PSLOHPTAPGMPPTQPPAPPTQ-PSPTPVSSGQT 899
OY 64 PDP 66
DB 900 PTP 902

RESULT 7
US-08-961-739-2
Sequence 2, Application US/08961739A
Patent No. 6063583
GENERAL INFORMATION:
APPLICANT: Montminy, Marc R.
TITLE OF INVENTION: Methods for Treating Diabetes Mellitus
FILE REFERENCE: SALK1650-1
CURRENT APPLICATION NUMBER: US/08/961,739A
CURRENT FILING DATE: 1997-10-31
EARLIER APPLICATION NUMBER: US 194,468
EARLIER FILING DATE: 1994-02-10
NUMBER OF SEQ ID NOS: 4
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 2441
TYPE: PRT
ORGANISM: Mus
FEATURE:
NAME/KEY: VARIANT
LOCATION: (1)...(2441)
OTHER INFORMATION: Xaa = Any Amino Acid
US-08-961-739-2

Query Match 17.1%; Score 71.5; DB 3; Length 2441;
Best Local Similarity 36.5%; Pred. No. 10;
Matches 23; Conservative 4; Mismatches 33; Indels 3; Gaps 2;

OY 4 SLPPRAAVPVLRMQPGAHVLSFLRPSMDLVSAFSLPLAPLSPTSPVSVGNG 63
DB 843 SOLPCPPTVQSLPHPTPPASTAGM--PSLOHPTAPGMPPTQPPAPPTQ-PSPTPVSSGQT 899
OY 64 PDP 66
DB 900 PTP 902

RESULT 8
US-08-615-170-21
Sequence 21, Application US/08615170
Patent No. 5776776
GENERAL INFORMATION:
APPLICANT: ORDAHL, Charles P.
APPLICANT: AZAKIE, Anthony
APPLICANT: MAR, Janet H.
APPLICANT: FARRANCE, Iain K.G.
APPLICANT: HALL, Deborah E.
APPLICANT: STEWART, Alexandre F.R.
TITLE OF INVENTION: DIEF-1 ISOFORMS AND USES THEREOF
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend Kourie and Crew
STREET: Steuart Street Tower, One Market Plaza
CITY: San Francisco
STATE: California
COUNTRY: US
ZIP: 94105-1493
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/615,170
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/01526
FILING DATE: 06-FEB-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/191,493
FILING DATE: 04-FEB-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Heslin, James M.
REGISTRATION NUMBER: 29,541
REFERENCE/DOCKET NUMBER: 23070-053120
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 326-2400
TELEFAX: (415) 326-2422
INFORMATION FOR SEQ ID NO: 21:
SEQUENCE CHARACTERISTICS:
LENGTH: 432 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-615-170-21

Query Match 16.6%; Score 69.5; DB 1; Length 432;
Best Local Similarity 35.0%; Pred. No. 2;
Matches 21; Conservative 11; Mismatches 17; Indels 11; Gaps 4;

OY 4 SLPPRAAVPVLRMQPGAHVLSFLRPSMDLVSAFSLPLAPLSPTSPVSVGNG 55
DB 155 SNAPEWGSGLP--GQPGSODIKRFAQPAVPDQPPMPPLASYS-PLAPLPAPASAVPV 211
RESULT 9
US-08-615-170-19
Sequence 19, Application US/08615170
Patent No. 5776776
GENERAL INFORMATION:
APPLICANT: ORDAHL, Charles P.
APPLICANT: AZAKIE, Anthony
APPLICANT: MAR, Janet H.
APPLICANT: FARRANCE, Iain K.G.

```
; APPLICANT: HALL, Deborah E.
; APPLICANT: STEWART, Alexandre F.R.
; APPLICANT: LARKIN, Sarah B.
; TITLE OF INVENTION: DREF-1 ISOFORMS AND USES THEREOF
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Kourie and Crew
; STREET: Steuart Street Tower, One Market Plaza
; CITY: San Francisco
; STATE: California
; COUNTRY: US
; ZIP: 94105-1493
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/615,170
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/01526
; FILING DATE: 06-FEB-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/191,493
; FILING DATE: 04-FEB-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Heslin, James M.
; REGISTRATION NUMBER: 29,541
; REFERENCE/DOCKET NUMBER: 23070-053120
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 326-2400
; TELEFAX: (415) 326-2422
; INFORMATION FOR SEQ ID NO: 19:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 433 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-615-170-19

Query Match          16.6%; Score 69.5; DB 1; Length 433;
Best Local Similarity 35.0%; Pred. No. 2;
Matches 21; Conservative 11; Mismatches 17; Indels 11; Gaps 4;

OY 4 SLRPPRAVPPLRMOPPAHVLSFLRPSMDVLS-----AFYSLPLAPLSP--TSVPI 55
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 156 SAAPRFWGPPI--GQPPSQDIKFAQAPVPIOPMPPLSLAYE-PLAPLPAPASAVPV 212

RESULT 10
; Sequence 58, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
```

```
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 115
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-697-58

Query Match          16.4%; Score 68.5; DB 4; Length 115;
Best Local Similarity 27.8%; Pred. No. 0.5;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPPRAVPVPLRMOP-----GPAH--PVLSFLRPSMDL-----VSAFY 41
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 11 GCHCMMSRDLTPPLPHEPSEGVLDCLGFCCHLLPLSPGSCWVLGHLRSLRPPSAASH 70

OY 42 SLPLAPLSPTSVPISPVS-----VGRG 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71 ALTISSLPPGLLPFGVGLTAHPQALIGRG 100

RESULT 11
; Sequence 54, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
; APPLICANT: Katz, Lawrence C.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
; TITLE OF INVENTION: CELL DEATH
; FILE REFERENCE: 10001-005-999
; CURRENT APPLICATION NUMBER: US/09/461,697
; CURRENT FILING DATE: 1999-12-14
; NUMBER OF SEQ ID NOS: 466
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 54
; LENGTH: 132
; TYPE: PRT
; ORGANISM: Homo sapiens
; US-09-461-697-54

Query Match          16.4%; Score 68.5; DB 4; Length 132;
Best Local Similarity 27.8%; Pred. No. 0.59;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPPRAVPVPLRMOP-----GPAH--PVLSFLRPSMDL-----VSAFY 41
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28 GCHCMMSRDLTPPLPHEPSEGVLDCLGFCCHLLPLSPGSCWVLGHLRSLRPPSAASH 87

OY 42 SLPLAPLSPTSVPISPVS-----VGRG 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 ALTISSLPPGLLPFGVGLTAHPQALIGRG 117

RESULT 12
; Sequence 52, Application US/09461697
; Patent No. 6277974
; GENERAL INFORMATION:
; APPLICANT: COGENT NEUROSCIENCE, Inc.
; APPLICANT: Lo, Donald C.
; APPLICANT: Barney, Shawn
; APPLICANT: Thomas, Mary Beth
; APPLICANT: Portbury, Stuart D.
; APPLICANT: Putnam, Kasturi
```

```

: APPLICANT: Katz, Lawrence C.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
: TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
: TITLE OF INVENTION: CELL DEATH
: FILE REFERENCE: 10001-005-999
: CURRENT APPLICATION NUMBER: US/09/461,697
: CURRENT FILING DATE: 1999-12-14
: NUMBER OF SEQ ID NOS: 466
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 52
: LENGTH: 139
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-461-697-52

Query Match          16.4%; Score 68.5; DB 4; Length 139;
Best Local Similarity 27.8%; Pred. No. 0.63;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPRAAVPVLKMP-----GPAH--PVLSFLRPSMDL-----VSARY 41
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 35 GCHCMSRDLTPLPHSPSPGVLDCLGPHLLPLSPGSPCWVLGHLHSPSAASH 94
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 42 SLPLAPLSPVSPISPV-----VGRG 63
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 95 ALTITSLPPGLLPFGVELTAHPQALIGRG 124

RESULT 13
US-09-461-697-48
: Sequence 48, Application US/09461697
: Patent No. 6277974
: GENERAL INFORMATION:
: APPLICANT: COGENT NEUROSCIENCE, Inc.
: APPLICANT: LO, Donald C.
: APPLICANT: Barney, Shawn
: APPLICANT: Thomas, Mary Beth
: APPLICANT: Portbury, Stuart D.
: APPLICANT: Puranam, Kasturi
: APPLICANT: Katz, Lawrence C.
: TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DIAGNOSING
: TITLE OF INVENTION: AND TREATING CONDITIONS, DISORDERS, OR DISEASES INVOLVING
: FILE REFERENCE: 10001-005-999
: CURRENT APPLICATION NUMBER: US/09/461,697
: CURRENT FILING DATE: 1999-12-14
: NUMBER OF SEQ ID NOS: 466
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 48
: LENGTH: 159
: TYPE: PRT
: ORGANISM: Homo sapiens
: US-09-461-697-48

Query Match          16.4%; Score 68.5; DB 4; Length 159;
Best Local Similarity 27.8%; Pred. No. 0.74;
Matches 25; Conservative 9; Mismatches 29; Indels 27; Gaps 4;

OY 1 GTHSLPRAAVPVLKMP-----GPAH--PVLSFLRPSMDL-----VSARY 41
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 55 GCHCMSRDLTPLPHSPSPGVLDCLGPHLLPLSPGSPCWVLGHLHSPSAASH 114
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 42 SLPLAPLSPVSPISPV-----VGRG 63
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 115 ALTITSLPPGLLPFGVELTAHPQALIGRG 144

RESULT 14
US-08-840-316-3
: Sequence 3, Application US/08840316
: Patent No. 6054567
```

```

: GENERAL INFORMATION:
: APPLICANT: Emerson, Suzanne U., Purcell, Robert H.,
: APPLICANT: Tsarev, Sergei, A., and Robinson, Robin A.
: TITLE OF INVENTION: Recombinant Proteins Of
: TITLE OF INVENTION: A Pakistani Strain Of Hepatitis E And Their
: TITLE OF INVENTION: Use In Diagnostic Methods And Vaccines
: NUMBER OF SEQUENCES: 111
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: MORGAN & FINNEGAN
: STREET: 345 PARK AVENUE
: CITY: NEW YORK
: STATE: NEW YORK
: COUNTRY: USA
: ZIP: 10154
: COMPUTER READABLE FORM:
: MEDIUM TYPE: FLOPPY DISK
: COMPUTER: IBM PC COMPATIBLE
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: WORDPERFECT 5.1
: CURRENT APPLICATION DATA:
: FILING DATE: 11-APR-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Richard W. Bork
: REGISTRATION NUMBER: 36,459
: REFERENCE/DOCKET NUMBER: 2026-4255
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (212) 758-4800
: TELEFAX: (212) 751-6849
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 123 amino acid residues
: TYPE: amino acid
: STRANDEDNESS: unknown
: TOPOLOGY: unknown
: US-08-840-316-3

Query Match          16.1%; Score 67.5; DB 3; Length 123;
Best Local Similarity 34.3%; Pred. No. 0.71;
Matches 23; Conservative 8; Mismatches 17; Indels 19; Gaps 3;

OY 10 AAVPV-----PLRMQGPAAHVLFLRPSMDLVA--FYSLLAPLSPT 51
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 48 AAVPAVSGVTGLILSPSQSPFIQPTPS-PMSPDLRPGLDLVFANPPDHSAPLGVTSPS 106
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
OY 52 SVSPISPV 58
    : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 107 APPLPHV 113

RESULT 15
US-08-478-507-9
: Sequence 9, Application US/08478507
: Patent No. 6120988
: GENERAL INFORMATION:
: APPLICANT: Reyes, Gregory R
: APPLICANT: Yarbough, Patrice O
: APPLICANT: Bradley, Daniel W
: APPLICANT: Krawczynski, Krzysztof Z
: APPLICANT: Tam, Albert
: APPLICANT: Fry, Kirk E
: TITLE OF INVENTION: DNA Sequences of Enterically Transmitted
: TITLE OF INVENTION: No. 6120988-A/No. 6120988-B Hepatitis Viral Agent
: NUMBER OF SEQUENCES: 20
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dehlinger & Associates
: STREET: 350 Cambridge Avenue, Suite 250
```



```

: CITY: Palo Alto
: STATE: CA
: COUNTRY: USA
: ZIP: 94306
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/478,507
: FILING DATE: 07-JUN-1995
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/279,823
: FILING DATE: 25-JUL-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/681,078
: FILING DATE: 05-APR-1991
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/505,888
: FILING DATE: 05-APR-1990
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/420,921
: FILING DATE: 13-OCT-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/367,486
: FILING DATE: 16-JUN-1989
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/336,672
: FILING DATE: 11-APR-1989
: APPLICATION NUMBER: US 07/208,997
: FILING DATE: 17-JUN-1988
: ATTORNEY/AGENT INFORMATION:
: NAME: Sholtz, Charles K.
: REGISTRATION NUMBER: 38,615
: REFERENCE/DOCKET NUMBER: 4600-0183.22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (650) 324-0880
: TELEFAX: (650) 324-0960
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 123 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-478-507-9

```

Query Match 16.1%; Score 67.5; DB 3; Length 123;
Best Local Similarity 34.3%; Pred. No. 0.71; Mismatches 17; Indels 19; Gaps 3;
Matches 23; Conservative 8;

QY 10 AAVPV-----PLRMQGPAPHPVLSFLRPSWDIVSA---FYSLEPLAPLSPT 51
||| | :||| :| :||| ||| | :||| | :
Db 48 AAVPAVAVSGVTGLISPSQSPIFIPTPS-PPMSPLRPGLDLVEANPDHSAFLGVTTRPS 106
QY 52 SVPISPV 58
:| :|
Db 107 APPLPHY 113

Search completed: April 11, 2002, 09:29:35
Job time: 72 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:08 ; Search time 40.19 seconds

(without alignments)
772.250 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287

Sequence: 1 MELALCRRGGLLALLPGEA.....VGRGPDPAHVAVNLSRYEG 419

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SIDS8/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SIDS8/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SIDS8/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SIDS8/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SIDS8/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SIDS8/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SIDS8/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SIDS8/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SIDS8/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SIDS8/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SIDS8/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SIDS8/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SIDS8/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SIDS8/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SIDS8/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SIDS8/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SIDS8/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SIDS8/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SIDS8/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SIDS8/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SIDS8/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SIDS8/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2158.5	94.4	420	AA97240	Truncated HER-2, p
2	1878	82.1	645	AA860408	Human ErbB2 oncopr
3	1878	82.1	645	AA861593	Human ErbB2 extrac
4	1878	82.1	653	AA821200	Extracellular HER-
5	1878	82.1	712	AA821204	Human HER-2/neu fu
6	1878	82.1	782	AA819764	HER-2/neu C-SF immu
7	1878	82.1	919	AA821203	Human HER-2/neu fu
8	1878	82.1	1200	AA821208	Human HER-2/neu pr
9	1878	82.1	1255	AA801111	HER-2/neu protein.
10	1878	82.1	1255	AA92406	Human HER-2/neu on
11	1878	82.1	1255	AA821198	Human HER-2/neu pr

12	1878	82.1	1255	AA84780	Amino acid sequenc
13	1878	82.1	1255	AA92620	Human heregulin 2
14	1878	82.1	1255	AA85458	Human HER-2/neu pr
15	1878	82.1	1255	AA888267	HER2/neu amino aci
16	1878	82.1	1255	AA860167	HER2 transgene pla
17	1860	81.3	1433	AA839568	Sequence of c-erbB
18	1775	77.6	951	AA44993	DC8scFv-erbB2EC fu
19	1736	75.9	624	AA808222	Extracellular port
20	1608.5	70.3	654	AA821205	Rat HER-2/neu prot
21	1608.5	70.3	1256	AA821199	Rat HER-2/neu prot
22	1597.5	69.9	1256	AA821206	Mouse Her-2/neu pr
23	1597.5	69.9	1256	AA821206	Amino acid sequenc
24	987	43.2	191	AA859354	Human ErbB2 protei
25	987	43.2	191	AA859354	Human ErbB2 recept
26	987	43.2	191	AA848767	Human ErbB2 N-term
27	860	37.6	166	AA859345	Human ErbB2 domain
28	860	37.6	166	AA862074	ErbB2 protein doma
29	860	37.6	166	AA848763	Human ErbB2 extrac
30	793	34.7	405	AA833737	Epidermal growth f
31	793	34.7	529	AA838209	Epidermal growth f
32	793	34.7	1210	AA819259	LD1D2D3.Appl EGF r
33	793	34.7	1210	AA850616	Amino acid sequenc
34	793	34.7	1210	AA850616	Human EGF receptor
35	782	34.2	621	AA868420	Amino acid sequenc
36	775	33.9	910	AA891737	Epidermal growth f
37	775	33.9	1058	AA854843	HER4-ig fusion pro
38	775	33.9	1058	AA891734	HER4 with alternat
39	775	33.9	1292	AA891734	Receptor tyrosine
40	775	33.9	1292	AA891734	Human epidermal gr
41	775	33.9	1308	AA854841	Human epidermal gr
42	775	33.9	1308	AA891733	HER4. Homo sapien
43	775	33.9	1308	AA864499	Receptor tyrosine
44	773	33.8	478	AA848033	Human inducible ni
45	735.5	32.2	1342	AA88453	EGF receptor relat

ALIGNMENTS

RESULT 1	
AA97240	AA97240 standard; protein; 420 AA.
XX	
AC	AA97240:
XX	
DT	04-DEC-2000 (first entry)
XX	
DE	Truncated HER-2, p68-HER-2.
XX	
KW	HER-2, erbB-2; oncogene; receptor-like tyrosine kinase; insertion;
KW	extracellular domain IIIa; antagonist; intron 8; C-terminal extension;
KW	truncated HER-2; p68; dimerization inhibitor; cytosolic.
XX	
OS	Homo sapiens.
XX	
XX	
Key	Location/Qualifiers
FT	Misc-difference 124
FT	/note= "represented as Agn"
FT	Misc-difference 125
FT	/note= "represented as Agn"
FT	Misc-difference 343
FT	/note= "Preferably Ser"
FT	Misc-difference 346
FT	/note= "Preferably Pro"
FT	Misc-difference 347
FT	/note= "Preferably Leu"
FT	Misc-difference 357
FT	/note= "Preferably Gln"
FT	Misc-difference 359
FT	/note= "Preferably Leu"
FT	Misc-difference 362
FT	/note= "changes from glycine"
FT	Misc-difference 377

```
FT /note= "Preferably Ile"
FT Misc-difference 395
FT /note= "Preferably Arg"
FT Misc-difference 405
FT /note= "Preferably Leu"
FT Misc-difference 414
FT /note= "Preferably Asn"
PN W0200044403-A1.
PD 03-AUG-2000.
PF 20-JAN-2000; 2000WO-US01484.
PR 20-JAN-1999; 99US-0234208.
PA (UYOR-) UNIV OREGON HEALTH SCT.
PI Doherty JK, Clinton GM, Adelman JP;
PS WPI; 2000-499287/44.
PT Using polypeptides and antibodies that bind to the extracellular domain
PT of the receptor-like tyrosine kinase HER-2 to treat solid tumors of the
PT breast, lung, ovaries and colon
PS Claim 8; Page 39-40; 46pp; English.
XX HER-2/neu (erbB-2) oncogene encodes a receptor-like tyrosine kinase. The
XX extracellular domain of p185-HER-2 is proteolytically shed from breast
XX carcinoma cells in culture and is found in serum of some cancer patients
XX and may be a serum marker of metastatic breast cancer. An alternative
XX HER-2 mRNA of 4.8 kb with a 274 bp insert (intron 8) has been
XX identified. The retained intron is in-frame and encodes a 79 amino acid
XX extension designated ECDIIa (the present sequence), which is inserted at
XX residue 340 of p185-HER-2. The alternative mRNA predicts a truncated
XX HER-2 protein (approximately 68 kDa) that lacks the transmembrane and
XX intracellular domains (see AAY97240). p68HER-2 specifically binds to
XX p185-HER-2 without activating HER-2. It could therefore block
XX dimerization of p185-HER-2. The p68HER-2 polypeptide binds to a site on
XX the ECD of HER-2 that is different from the site of binding for
XX Hereceptin (Rm) (a marketed humanized monoclonal antibody that is used
XX for the treatment of cancer and binds to the ECD of HER-2). The methods,
XX compositions, polypeptides and antibodies are used to treat solid
XX tumours such as breast cancer, small cell lung carcinoma, ovarian cancer
XX and/or colon cancer, especially where over-expression of HER-2 is
XX indicated.
SQ Sequence 420 AA:
Query Match 94.4%; Score 2158.5; DB 21; Length 420;
Best Local Similarity 95.2%; Pred. No. 2,1e-166;
Matches 400; Conservative 0; Mismatches 19; Indels 1; Gaps 1;
QY 1 METALALCRNGILLALLPQCAASTQVCTGDMKLRLPASPEFHLMKRLHLYGCGQVVGNL 60
DB 1 melaaletcwgllalallppgaastqvtctdcklrlpaspehldmlrlhlygcyvvgnl 60
QY 61 EETVPTNASLSFLDDIOEVGCVYLIANNOVQVPLQRRIRYRGVQLPEPDNALVLNDG 120
DB 61 eltylptnaslsflqddievvgvyliahnqvrvpdlqrlrlyrgvqllednyalavldng 120
QY 121 DLUNNTPTVPGASPGCLRELQRLSLTEILKGCVLQIQRNPQLCYOPTILMKDIFHHNNOLA 180
DB 121 dlunntptvpgaspgclrelqlslteilkgcvglqiornpqlcyoptilmkdifhnnola 180
QY 181 LTLIDPNSRACHPCSPCKGSRGCESESDQSLRTVCAGACARCKGPLPTDCCHQC 240
DB 181 ltlidpnsrachpcspckgsrcgcesesdqslrtvcagacarckgplptdcchqc 240
QY 241 AAGCTGPKHSDCLACLHNHSGICELHCPALVTYNTDFESHPNBEGRITFASCVTACP 300
DB 241 aagctgpkhsdclaclhnhsgicelhcpalvtyn tdfeshpnbegr itfascvtacd 300
```

```
DB 241 aagctgpkhsdclaclhnhsgicelhcpalvtyn tdfescpnpegrylfigascvtacp 300
QY 301 YNYLSTDVSGCTLWCPLPHNOEVTAEDETQRCCKSKPCAR-GTHSLPLPPAAVPPPLRMQ 359
DB 301 ynlstdvsgctlvcp lphngetaedgtqrcckskpcarvgvnxhxxprpaavpppxrx 360
QY 360 PGEAHPVLSTLRPSWDLVSAFYSLPLAPLSPTSPVISPVYSGRPPDAVAANLSRYEG 419
DB 361 pxpahpvlstlrpswdlvsa fyslplaplsp tsvpvispvysgrppdahvaanlsryeg 420
RESULT 2
AAB60408
ID AAB60408 standard; Protein; 645 AA.
AC AAB60408;
DT 24-APR-2001 (first entry)
DE Human ErbB2 oncoprotein, SEQ ID NO:13.
XX Anti-ErbB2 monoclonal antibody 2C4; HER2; mouse; murine; humanised; VL;
XX light chain variable region; cancer; cytostatic; EGFR-expressing cancer;
XX epidermal growth factor receptor; colon cancer; rectal cancer; tumour;
XX colorectal cancer; non-small cell lung cancer; metastatic breast cancer;
XX affinity purification.
XX Homo sapiens.
XX OS
XX PN W0200100245-A2.
XX PD 04-JAN-2001.
XX PF 23-JUN-2000; 2000WO-US17366.
XX PR 25-JUN-1999; 99US-0141316.
XX PA (GETH ) GENENTECH INC.
XX PI Adams CW, Presta LG, Sliwkowsky M;
XX WPI; 2001-080862/09.
XX DR Treating cancer in a human, where the cancer expresses epidermal growth
XX factor receptor (EGFR), comprises administering an antibody which binds
XX to ErbB2 -
XX PS Example 1; Fig 1A; 89pp; English.
XX The invention relates to a method for treating cancer in a human patient,
XX wherein the cancer expresses epidermal growth factor receptor (EGFR),
XX comprising administering an antibody which binds ErbB2 (HER2; AAB60408).
XX In particular, the anti-ErbB2 antibody is the murine monoclonal antibody
XX 2C4 (AAB60396, AAB60397) or a humanised version of 2C4 (AAB60398,
XX AAB60399). The invention also encompasses an isolated nucleic acid
XX encoding a humanised ErbB2-binding antibody; vectors and host cells
XX comprising such nucleic acids; the recombinant production of a humanised
XX ErbB2-binding antibody; and an immunoconyugate comprising a humanised
XX ErbB2-binding antibody and a cytotoxic drug. The ErbB2-binding antibodies
XX act by antagonising ErbB receptors, and as inhibitors of transforming
XX growth factor alpha (TGF-alpha)-activated mitogen activated protein
XX kinase (MAPK). The method of the invention is used for treating cancer,
XX especially colon cancer, rectal cancer, colorectal cancer, lung cancer
XX (especially non-small cell lung cancer), or breast cancer (especially
XX metastatic breast cancer). The antibodies may also have non-therapeutic
XX uses e.g., as affinity purification agents. Using an antibody which binds
XX to ErbB2 to treat cancer is preferable to the use of EGFR-targeted
XX drugs, as EGFR is also highly expressed in other tissues such as the
XX liver and skin, where the active drug will also bind, with skin toxicity
XX having been observed for EGFR-targeted drugs. Antibodies which bind
XX ErbB2 are anticipated to have a better safety profile than such drugs.
XX The present sequence represents human ErbB2.
```

Sequence 645 AA:

Query Match 82.1%; Score 1878; DB 22; Length 645;
Best Local Similarity 83.0%; Pred. No. 1.7e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

1 MELALCRWGLLALLPRGASTQVCTGTDMLRLPASPETHLMDLRHLVGGCVVQGNL 60
1 melalcrwgl1allppgaastqvcitgtdmklrlpaspehldmlrhllygqcvvqgnl 60
61 ELTYLPTNASLSFLDIOEVQGYVLIANNQVROVPLQRLRIYRGTOLEPDNYALAVLNG 120
61 eltylptnaslsflqdiqevgyvliahnqvrvplqrlriyrgtqlfednyalavldng 120
121 DPLNNTPTVTGASPGGLRELDLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
121 dplnntptvtgaspzglrelqrlsteilkggvliqrnpqlcyqdtllwkdifhknq1a 180
181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLPTDCCHQC 240
181 ltlidtnsrachpcspmccksgrcwgessecdqsltrtcaggarckgplptdcchqc 240
241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
241 aagctgpkhsdclac1fhnsigicelhcpalvtynldfesmnpnegryltfgascvtacp 300
301 YNYLSTDVSGCTLVCPLENNQEVTAEDGTQRCCKSKPCAR-----GTHSLPRPAAVPVP 355
301 ynylstdvsgctlvcp1nnqevtaedgtqrcckskpcarvcyglgmehlrevravtsan 360
356 LRMQCG--PAHPVLSFLRPMNDVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
356 lrmqcg--pahpvlslflrpsmdlvsaflslplaplsptsvpi-----spvsvgrgpd 405
406 --PDAHVAVNLSRYEG 419
406 --pdahvaavnlsryeg 419
418 slpdlsvfqnlgvltg 433
418 slpdlsvfqnlgvltg 433

RESULT 3
AAB61593
ID AAB61593 standard; protein; 645 AA.
XX
AC AAB61593;
XX
DT 04-APR-2001 (first entry)
XX
DE Human ErbB2 extracellular domain.
XX
KW Human: ErbB2; cytosolic; prostate cancer; receptor tyrosine kinase;
KW antibody; ErbB receptor; monoclonal antibody 2C4; variable light chain.
XX
OS Homo sapiens.
XX
PN W0200100238-A1.
XX
PD 04-JAN-2001.
XX
PE 23-JUN-2000; 2000WO-US17423.
XX
PR 25-JUN-1999; 99US-0141315.
XX
PA (GETH.) GENENTECH INC.
PA (SLOK.) SLOAN KETTERING INST CANCER RES.
XX
PI Agus DB, Scher HI, Sliwkowski MX;
XX
DR WPI; 2001-159131/16.
XX
PT Treating prostate cancer in a human comprises administering an antibody
PT which binds ErbB2 and blocks ligand activation of an ErbB receptor -
XX

PS Disclosure; Fig 1; 93pp; English.

XX The ErbB family of receptor tyrosine kinases are important mediators of
CC cell growth, differentiation and survival. The receptor family includes
CC four distinct members including Epidermal Growth Factor Receptor (EGFR or
CC ErbB1), HER2 (ErbB2 or p185^{neu}), HER3 (ErbB3) and Her4 (ErbB4 or Tyro2).
CC The present invention relates to a method for treating prostate cancer.
CC The method comprises administering an antibody which binds ErbB2 and
CC blocks ligand activation of an ErbB receptor. Preferably, the antibody
CC blocks binding of monoclonal antibody 2C4 to ErbB2 and/or blocks
CC TGF-alpha activation of mitogen-activated protein kinase (MAPK). The
CC present sequence is the extracellular domain of human ErbB2.
XX

Sequence 645 AA:

Query Match 82.1%; Score 1878; DB 22; Length 645;
Best Local Similarity 83.0%; Pred. No. 1.7e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

1 MELALCRWGLLALLPRGASTQVCTGTDMLRLPASPETHLMDLRHLVGGCVVQGNL 60
1 melalcrwgl1allppgaastqvcitgtdmklrlpaspehldmlrhllygqcvvqgnl 60
61 ELTYLPTNASLSFLDIOEVQGYVLIANNQVROVPLQRLRIYRGTOLEPDNYALAVLNG 120
61 eltylptnaslsflqdiqevgyvliahnqvrvplqrlriyrgtqlfednyalavldng 120
121 DPLNNTPTVTGASPGGLRELDLRSLTEILKGGVLIQRNPOLCYODTILMKDIFHKNNOLA 180
121 dplnntptvtgaspzglrelqrlsteilkggvliqrnpqlcyqdtllwkdifhknq1a 180
181 LTLIDTNRSRACHPCSPMCKSGRCWGESSEDCQSLTRTVACAGCARCKGPLPTDCCHQC 240
181 ltlidtnsrachpcspmccksgrcwgessecdqsltrtcaggarckgplptdcchqc 240
241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
241 aagctgpkhsdclac1fhnsigicelhcpalvtynldfesmnpnegryltfgascvtacp 300
301 YNYLSTDVSGCTLVCPLENNQEVTAEDGTQRCCKSKPCAR-----GTHSLPRPAAVPVP 355
301 ynylstdvsgctlvcp1nnqevtaedgtqrcckskpcarvcyglgmehlrevravtsan 360
356 LRMQCG--PAHPVLSFLRPMNDVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
356 lrmqcg--pahpvlslflrpsmdlvsaflslplaplsptsvpi-----spvsvgrgpd 405
406 --PDAHVAVNLSRYEG 419
406 --pdahvaavnlsryeg 419
418 slpdlsvfqnlgvltg 433
418 slpdlsvfqnlgvltg 433

RESULT 4
AAB21200
ID AAB21200 standard; protein; 653 AA.
XX
AC AAB21200;
XX
DT 12-JAN-2001 (first entry)
XX
DE Extracellular HER-2/neu protein.
XX
KW HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer.
XX
OS unidentified.
XX
PN W0200044899-A1.
XX
PD 03-AUG-2000.
XX

PF 28-JAN-2000; 2000WO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.
 XX
 PA (CORI-) CORIXA CORP.
 XX (SMIK) SMITHLINE BEECHAM.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI: 2000-505976/45.
 XX
 PR HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX
 PS Claim 2: Fig 9; 128pp; English.
 XX
 CC The present sequence is the extracellular HER-2/neu protein. HER-2/neu is
 CC a member of the tyrosine kinase family of receptor-like glycoproteins and
 CC shows homology to the epidermal growth factor receptor (EGFR). It
 CC probably plays a part in cell growth and/or differentiation. The
 CC HER-2/neu gene is an oncogene. An HER-2/neu fusion protein comprising a
 CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
 CC domain may be used to treat or prevent cancer by eliciting or
 CC enhancing an immune response to the HER-2/neu protein. It may be used
 CC to treat malignancies such as breast, ovarian, colon, lung and
 CC prostate cancers, and may be used as an antigen to vaccinate against
 CC these neoplasias.
 XX
 SQ Sequence 653 AA;

Query Match 82.1%; Score 1878; DB 21: Length 653;
 Best Local Similarity 83.0%; Pred. No. 1.7e-143;

Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 MELAALCRWGLLALLPGCASTOVCTGDMKRLRPASPEHLDMLRHLRYGCGVVOGNTL 60
 DB 1 melaalcrwglallallpgcaastyctgdmkrlrpaspehldmlrhlrygscvvgvgnl 60
 OY 61 ELTYPTNASTSLFDIOIOGVGYLIAHNOVQVPLQRLRIYRGVTOLEFEDNALAVLDNG 120
 DB 61 eltyptnastslfdioiovgvyliahnovqvplqrlriyrgvtqlfednyalavldng 120
 OY 121 DLNMTPTVTGASPGGLRELQRLSLTEILKGVLIQRIORNPOLCYODTIIMKDFHKNNOA 180
 DB 121 dlnmtptvtgaspgglrelqrlslteilkgvliqriornpolcyodtiimkdfhknnga 180
 OY 181 LTLIDTNSRACHPCSPMCKGSRGWESSEDCQSLRTVACAGCARCKGPLPTDCHEQC 240
 DB 181 ltlidtnsrachpcspmkgsrwcgessedcqsllrtvcagcarckgplptdccheqc 240
 OY 241 AAGCTGPKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNREGRTFGASCVTACP 300
 DB 241 aagctgpkhsddlacilhfnhsgicelhcpalvtynntdtfesmpnregrytfgascvtacp 300
 OY 301 YVYLDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPP 355
 DB 301 yvyldvgsctlvcpplhnoevtaeedgtorcckskpcar-----gthslrppaavpp 355
 OY 356 LRMOPG--PAHVLSPFLRSPMDLVSAFYSLPAPLSPVPI-----SPVSVCGRPD 405
 DB 356 lrmopg--pahvlsplflrspmdlvsaftyslpaplspvpi-----spvsvcgrpd 405
 OY 406 --PDAAVAVNLSTRYEG 419
 DB 406 --pdavavnlstryeg 419
 OY 418 sipdlsvfqnlgvixg 433
 DB 418 sipdlsvfqnlgvixg 433

RESULT 5
 AAB21204
 ID AAB21204 standard; protein; 712 AA.
 XX

AC AAB21204;
 XX
 DT 12-JAN-2001 (first entry)
 XX
 DE Human HER-2/neu fusion protein.
 XX
 KW Human; HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;
 KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
 KW colon cancer; fusion protein.
 XX
 OS Homo sapiens.
 OS Synthetic.
 XX
 PN WO200044899-A1.
 XX
 PD 03-AUG-2000.
 XX
 PF 28-JAN-2000; 2000WO-US02164.
 XX
 PR 29-JAN-1999; 99US-0117976.
 XX
 PA (CORI-) CORIXA CORP.
 XX (SMIK) SMITHLINE BEECHAM.
 XX
 PI Cheever MA, Gheysen D;
 XX
 DR WPI: 2000-505976/45.
 XX
 PR HER-2/neu extracellular domain/phosphorylation domain fusion proteins
 PT useful for vaccinating against breast, ovarian, colon, lung and
 PT prostate cancers -
 XX
 PS Claim 27: Fig 13; 128pp; English.
 XX
 CC The present sequence is a fusion protein comprising the extracellular
 CC domain and a preferred portion of the phosphorylation domain of the human
 CC HER-2/neu protein. HER-2/neu is a member of the tyrosine kinase family of
 CC receptor-like glycoproteins and shows homology to the epidermal growth
 CC factor receptor (EGFR). It probably plays a part in cell growth and/or
 CC differentiation. The HER-2/neu gene is an oncogene. HER-2/neu fusion
 CC proteins may be used to treat or prevent cancer by eliciting or enhancing
 CC an immune response to the HER-2/neu protein. They may be used to treat
 CC malignancies such as breast, ovarian, colon, lung and prostate cancers,
 CC and may be used as an antigen to vaccinate against these neoplasias.
 XX
 SQ Sequence 712 AA;

Query Match 82.1%; Score 1878; DB 21: Length 712;
 Best Local Similarity 83.0%; Pred. No. 1.9e-143;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 MELAALCRWGLLALLPGCASTOVCTGDMKRLRPASPEHLDMLRHLRYGCGVVOGNTL 60
 DB 1 melaalcrwglallallpgcaastyctgdmkrlrpaspehldmlrhlrygscvvgvgnl 60
 OY 61 ELTYPTNASTSLFDIOIOGVGYLIAHNOVQVPLQRLRIYRGVTOLEFEDNALAVLDNG 120
 DB 61 eltyptnastslfdioiovgvyliahnovqvplqrlriyrgvtqlfednyalavldng 120
 OY 121 DLNMTPTVTGASPGGLRELQRLSLTEILKGVLIQRIORNPOLCYODTIIMKDFHKNNOA 180
 DB 121 dlnmtptvtgaspgglrelqrlslteilkgvliqriornpolcyodtiimkdfhknnga 180
 OY 181 LTLIDTNSRACHPCSPMCKGSRGWESSEDCQSLRTVACAGCARCKGPLPTDCHEQC 240
 DB 181 ltlidtnsrachpcspmkgsrwcgessedcqsllrtvcagcarckgplptdccheqc 240
 OY 241 AAGCTGPKHSDDLACILHFNHSGICELHCPALVTYNTDTFESMPNREGRTFGASCVTACP 300
 DB 241 aagctgpkhsddlacilhfnhsgicelhcpalvtynntdtfesmpnregrytfgascvtacp 300
 OY 301 YVYLDVGSCTLVCPPLHNOEVTAEEDGTORCEKSKPCAR-----GTHSLRPPAAVPP 355

|||||
Db 301 ynylstdvgsctlvcp|hngvtaedgtqrcekskpcarvcyglgmehllrevravtsan 360
QY 356 LRMQPG--PAHPVLSFLRSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 lgefagckkifgslafpsfdgdpasnt---ap|qpeq|ygfetleeltgy|lyisawpd 417
QY 406 --PDAHVAVNLSRREG 419
Db 418 slpdlsvfqn|qvtrg 433

RESULT 6
AAM19764
ID AAM19764 standard; Protein; 782 AA.
AC AAM19764;
XX
XX 17-SEP-1997 (first entry)
DE Her2-GM-CSF immunostimulant fusion protein.
XX
XX Her2-GM-CSF; granulocyte macrophage colony stimulating factor;
KW growth factor receptor; oncogene; immunostimulant; cancer;
KW therapy.
XX
XX Homo sapiens.
OS
FH Key Location/Qualifiers
FT Protein 1..653
FT /label= Her2
FT Peptide 654..655
FT /label= Linker
FT Protein 656..782
FT /label= GM-CSF
XX
XX WO9724438-A1.
XX
XX 10-JUL-1997.
PD
XX 23-DEC-1996; 96WO-US20241.
PF
XX 28-DEC-1995; 95US-0579823.
PR
XX (ACT1-) ACTIVATED CELL THERAPY INC.
PA
PI Laus R, Ruegg CL, Wu H;
XX
XX WPI: 1997-363674/33.
DR N-PSDB; AAT72725.
XX
XX Potent APC that activates T-cells to give multivalent cellular
PT immune response - can also induce a cytotoxic T-cell response in a
PT vertebrate subject
XX
XX Disclosure: Fig 8; 45pp; English.
XX
XX A fusion protein (AAM19764) comprises Her2 (a growth factor receptor
CC that is overexpressed in breast, ovarian can other cancer cells)
CC and granulocyte-macrophage colony stimulating factor (GM-CSF). It
CC is the expression product of a nucleic acid molecule (AAT72725)
CC prep'd. by PCR amplification of Her2 cDNA from a breast cancer cell
CC line and fusion to GM-CSF cDNA. Fusion expression vectors can be
CC used to transfect mammalian and insect cells. The Her2-GM-CSF
CC fusion protein is used to generate anti-Her2 immunity. Tumour
CC cells are eliminated by cytotoxic T lymphocytes activated in vivo
CC or in vitro by exposure to antigen-presenting cells exposed to the
CC fusion protein.
XX
XX Sequence 782 AA:
SQ

Query Match 82.1%; Score 1878; DB 18; Length 782;

Best Local Similarity 83.0%; Pred. No. 2,2e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRNGLLALLPPRAASTQVCTGDMKRLRPASETHLMDLRHYGOCQVVGWL 60
Db 1 melaa|crngll|allppraastqvc|tdmk|rlrpaspeth|dmlrhygqcvvgwnl 60
QY 61 E|T|Y|P|T|N|S|F|L|D|I|O|V|G|V|L|A|H|N|Q|V|R|P|L|R|Y|R|G|T|O|L|F|E|D|N|A|L|V|D|N|G 120
Db 61 e|t|y|p|t|n|s|f|l|d|i|o|v|g|v|l|h|n|q|v|r|p|l|r|y|r|g|t|o|l|f|e|d|n|a|l|v|d|n|g 120
QY 121 DPLNMTTPVYGASPGGLARELQRLSTELIKGGVL|Q|RN|P|Q|C|Y|O|D|T|I|K|D|I|F|H|K|N|N|D|L 180
Db 121 dplnmttpv|y|g|a|s|p|g|l|a|r|e|l|q|r|l|s|t|e|l|i|k|g|v|l|q|r|n|p|q|c|y|o|d|t|i|k|d|i|f|h|k|n|n|d|l 180
QY 181 L|T|I|D|T|R|N|R|R|A|C|H|P|C|S|P|W|K|G|R|C|W|G|E|S|E|D|Q|S|L|R|T|Y|C|A|G|C|A|R|C|K|P|L|T|D|C|H|E|Q 240
Db 181 l|t|i|d|t|r|n|r|r|a|c|h|p|c|s|p|w|k|g|r|c|w|g|e|s|e|d|q|s|l|r|t|y|c|a|g|c|a|r|c|k|p|l|t|d|c|h|e|q 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 aagctgpkh|s|d|c|l|a|c|h|f|n|h|s|g|i|c|e|h|c|p|a|l|v|t|y|n|t|d|f|e|s|m|p|n|d|e|g|r|y|t|f|g|a|s|c|v|t|a|c|p 300
QY 301 YNYLSTDVGSCTLVCP|LHNGEYTAEDGTQRCCKSPCAR-----GTHSLP|P|P|A|V|P|P 355
Db 301 ynylstdvgsctlvcp|hngvtaedgtqrcekskpcarvcyglgmehllrevravtsan 360
QY 356 LRMQPG--PAHPVLSFLRSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
Db 361 lgefagckkifgslafpsfdgdpasnt---ap|qpeq|ygfetleeltgy|lyisawpd 417
QY 406 --PDAHVAVNLSRREG 419
Db 418 slpdlsvfqn|qvtrg 433

RESULT 7
AAB21203
ID AAB21203 standard; protein; 919 AA.
XX
XX AAB21203;
AC
XX 12-JAN-2001 (first entry)
DE Human HER-2/neu fusion protein.
XX
XX Human; HER-2/neu; oncogene; tyrosine kinase; cytostatic; vaccine;
KW breast cancer; prostate cancer; ovarian cancer; lung cancer;
KW colon cancer; fusion protein.
XX
XX Homo sapiens.
OS
OS Synthetic.
PN WO200044899-A1.
PD
XX 03-AUG-2000.
XX
XX 28-JAN-2000; 2000WO-US02164.
PF
XX 29-JAN-1999; 99US-0117976.
PR
XX (CORI-) CORIXA CORP.
PA (SMK) SMITHKLINE BEECHAM.
XX
XX Cheever MA, Gheysen D;
PI
XX WPI: 2000-505976/45.
DR
XX
XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins
PT useful for vaccinating against breast, ovarian, colon, lung and
PT prostate cancers -
XX
XX Claim 2; Fig 12; 128pp; English.

XX The present sequence is a fusion protein comprising the extracellular
CC domain and the phosphorylation domain of the human HER-2/neu protein.
CC HER-2/neu is a member of the tyrosine kinase family of receptor-like
CC glycoproteins and shows homology to the epidermal growth factor receptor
CC (EGFR). It probably plays a part in cell growth and/or differentiation.
CC The HER-2/neu gene is an oncogene. HER-2/neu fusion proteins may be used
CC to treat or prevent cancer by eliciting or enhancing an immune response
CC to the HER-2/neu protein. They may be used to treat malignancies such as
CC breast, ovarian, colon, lung and prostate cancers, and may be used as an
CC antigen to vaccinate against these neoplasias.

XX Sequence 919 AA:

Query Match 82.1%; Score 1878; DB 21; Length 919;
Best Local Similarity 83.0%; Pred. No. 2.7e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METAAICRMGILLALIPGAASVOVCTGDMKRLRPASPEHLDMLRHLXOGCQVVGNTL 60
DB 1 metaaicrmgillalippgaastvcyctgdmkrlrpaspethldmlrhllygqcqvvgntl 60
QY 61 ELTYLPTNASLSFLDIOEVQGVYLIANQVNPQLQRLRIYRGTOLEFEDNTALAVLDNG 120
DB 61 eltylptnaslsflqdigevgyvlianhqvrvpqlrlrivrqtqlfednyalavldng 120
QY 121 DELNNTPTVTGASPGSLRQLRSLTEILKGGVLIQRNQOLCYODTILMKDIFHKNNOLA 180
DB 121 delnntptvtgaspzglrqlrsltelikgvlilqrnqolcyqdtllmkdihknnola 180
QY 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCOSLTRVCAGGACARCKGPLPTDCCHEQC 240
DB 181 ltlidntrsrachpcsmcksgrcwgessecdqslrtvcaggacarkgplptdccheqc 240
QY 241 AAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNBEGRYTFGASCVTACP 300
DB 241 aagctgpkhsdclacihfhnsigicelhcpalvtyntdfesmpnbeqyrtfgascvtacp 300
QY 301 YNYLSTDVGSCITLVCPILHNOEYTAEDGTORCEKSKPCAR-----GTHSLRPRAVPVP 355
DB 301 ynylstdvgsctlvcpilhnqevtaedgtqrcekskpcarvcyglgmehllrevratlsan 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSIFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 lrmqpg--pahpvlslflrpsmdlvsafyslplaplspstsvpi-----spvsvgrgpd 405
QY 406 igeifagckkifgsaflfpefdgdpasnt---apldpeqlqvretleeltyglylsawpd 417
DB 406 igeifagckkifgsaflfpefdgdpasnt---apldpeqlqvretleeltyglylsawpd 417
QY 418 slpdlsvfqnlgvrlg 433
DB 418 slpdlsvfqnlgvrlg 433

RESULT 8

AAB21208
ID AAB21208 standard; Protein; 1200 AA.

XX AAB21208;

DT 12-JAN-2001 (first entry)

XX Human HER-2/neu protein.

XX Human: HER-2/neu; oncogene; tyrosine kinase; cytosolic; vaccine;

KM breast cancer; prostate cancer; ovarian cancer; lung cancer;

XX colon cancer.

OS Homo sapiens.

XX MO200044899-A1.

XX 03-AUG-2000.

XX 28-JAN-2000; 2000WO-US02164.

XX 29-JAN-1999; 99US-0117976.

XX (CORI-) CORIAX CORP.

XX (SMK) SMITHKLINE BEECHAM.

XX Cheever MA, Gheysen D;

XX MPI: 2000-505976/45.

XX N-PSDB: AAA89736.

XX HER-2/neu extracellular domain/phosphorylation domain fusion proteins

XX useful for vaccinating against breast, ovarian, colon, lung and

XX prostate cancers -

XX Disclosure: Fig 15; 128pp; English.

CC The present sequence is the human HER-2/neu protein. It is a member
CC of the tyrosine kinase family of receptor-like glycoproteins and shows
CC homology to the epidermal growth factor receptor (EGFR). It probably
CC plays a part in cell growth and/or differentiation. The HER-2/neu
CC gene is an oncogene. An HER-2/neu fusion protein comprising a
CC HER-2/neu extracellular domain fused to a HER-2/neu phosphorylation
CC domain may be used to treat or prevent cancer by eliciting or
CC enhancing an immune response to the HER-2/neu protein. It may be used
CC to treat malignancies such as breast, ovarian, colon, lung and
CC prostate cancers, and may be used as an antigen to vaccinate against
CC these neoplasias.

XX Sequence 1200 AA:

Query Match 82.1%; Score 1878; DB 21; Length 1200;
Best Local Similarity 83.0%; Pred. No. 3.8e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METAAICRMGILLALIPGAASVOVCTGDMKRLRPASPEHLDMLRHLXOGCQVVGNTL 60
DB 1 metaaicrmgillalippgaastvcyctgdmkrlrpaspethldmlrhllygqcqvvgntl 60
QY 61 ELTYLPTNASLSFLDIOEVQGVYLIANQVNPQLQRLRIYRGTOLEFEDNTALAVLDNG 120
DB 61 eltylptnaslsflqdigevgyvlianhqvrvpqlrlrivrqtqlfednyalavldng 120
QY 121 DELNNTPTVTGASPGSLRQLRSLTEILKGGVLIQRNQOLCYODTILMKDIFHKNNOLA 180
DB 121 delnntptvtgaspzglrqlrsltelikgvlilqrnqolcyqdtllmkdihknnola 180
QY 181 LTLIDNTRSRACHPCSPMCKSGRCWGESSEDCOSLTRVCAGGACARCKGPLPTDCCHEQC 240
DB 181 ltlidntrsrachpcsmcksgrcwgessecdqslrtvcaggacarkgplptdccheqc 240
QY 241 AAGCTGPKHSDCLACIHFHNSGICELHCPALVTYNTDFESMPNBEGRYTFGASCVTACP 300
DB 241 aagctgpkhsdclacihfhnsigicelhcpalvtyntdfesmpnbeqyrtfgascvtacp 300
QY 301 YNYLSTDVGSCITLVCPILHNOEYTAEDGTORCEKSKPCAR-----GTHSLRPRAVPVP 355
DB 301 ynylstdvgsctlvcpilhnqevtaedgtqrcekskpcarvcyglgmehllrevratlsan 360
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSIFYSLPLAPLSPTSVPI-----SPVSVGRGPD 405
DB 356 lrmqpg--pahpvlslflrpsmdlvsafyslplaplspstsvpi-----spvsvgrgpd 405
QY 406 igeifagckkifgsaflfpefdgdpasnt---apldpeqlqvretleeltyglylsawpd 417
DB 406 igeifagckkifgsaflfpefdgdpasnt---apldpeqlqvretleeltyglylsawpd 417
QY 418 slpdlsvfqnlgvrlg 433
DB 418 slpdlsvfqnlgvrlg 433

RESULT 9

AAW01111
ID AAW01111 standard; Protein; 1255 AA.

XX

XX	AAW01111;	
XX		
DT	01-JAN-1997	(first entry)
XX		
DE	HER-2/neu protein.	
XX		
KW	HER-2/neu; c-erbB1; p185; oncogene; tyrosine protein kinase;	
KW	breast cancer; ovary cancer; colon cancer; lung cancer;	
XX	prostate cancer; immunisation; tumour; vaccine; vector.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Domain	676..1255
FT		/label="Intracellular-domain
FT		/note="Claimed domain, useful for immunisation"
XX		
PN	W09630514-A1.	
XX		
XX	03-OCT-1996.	
XX		
PF	28-MAR-1996;	96WO-US01689.
XX		
PR	31-MAR-1995;	95US-0414417.
XX		
PA	(UNIV) UNIV WASHINGTON.	
XX		
PI	Cheever MA, Disis ML;	
DR	WPI: 1996-455361/45.	
DR	N-PSDB; AAT40739.	
XX		
PT	DNA encoding HER-2-neu polypeptide(s) - used for prevention or	
PT	treatment of malignancies with which the HER-2/neu oncogene is	
PT	associated	
XX		
PS	Claim 2; Page 56-61; 71pp; English.	
XX		
XX	Human HER-2/neu protein (AAW01111), also called p185 or c-erbB2, is the product of the HER-2/neu oncogene (see also AAT40739). The protein is over-expressed in various cancers, including breast, ovarian, colon, lung and prostate. The intracellular domain of the protein can be used to immunise an animal against a malignancy with which the oncogene is associated. The polypeptide can be produced in transformed host cells for use in immunisation. Alternatively, animal cells are transfected in vivo or ex vivo with a viral vector that directs expression of the polypeptide.	
XX	Sequence	1255 AA;
XX		

Query Match	82.1%	Score 1878	DB 17	Length 1255
Best Local Similarity	83.0%	Pred. No. 4e-143		
Matches 362	Conservative 9	Mismatches 45	Indels 20	Gaps
QY	1	MELALCRWGLLALPRGASTVCTGTDKILPASPETHIDMLRHLYOGCVOGNTL	60	
Db	1	melaalcrwgljllalppgaastqvcyctdmkrlrpsapethidmlrhlyqgcvgvgnl	60	
QY	61	ELTLPTNASLSFLQDIOEVGVYVLIAHNOVROPLORLRIWSTQLFEDNYALAVLDG	120	
Db	61	elylptnaslsifldiqdevgyvllahngvryplqrlrlwvrtqglfednyalavldng	120	
QY	121	DPLNNTPTVTSAGSGRELDLRSLTFETIKGGVLIQNPOLCYODTILMDIPIKNNOLA	180	
Db	121	dplmntprvtsagsgjreldlrsltfetjqlrjlsrtellxygvlrlqmpqlcyqddlllwkdilthknqja	180	
QY	181	LTLIDTNKSRACHCSPCKGSRMGESSIEDCOSILRTTVCAGGACARKGRLPTDCCHEOC	240	
Db	181	ltlidtnsrachcspmpckgsrwmgsesedcglrlttvcaggarckgrrlpdcccnegc	240	
QY	241	AAGCTGPRHSCLACLFPHNSGICELHCPALVYNTDTFESMPNBERYTFGASCVTACP	300	

```

Db      241 aagctgphkhsclac\hfnhsigcel\hnpalvtn\tdtfesmpnpegryltfgascvaccp 3000
Qy      301 YNLTSDVSGCTLVCPLHNDVETAEDGTQRCEKCKSPCAR----GTHSLLPRAAPVVP 3555
        |||||
Db      301 ynlstdvsgctllvcplhngvetaedgtqrceckspcarvcyglgmeh\lrevrvtlsan 3600
        |||||
Qy      336 LRMDPG--PAMPVLSFLRPSMDVLSATYSLPLAPLSTSVPI-----SPVSVGCGPD 4050
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      361 igefagckktifgsiaflpessfdgpnasnt--aplqpegqvlfetleelctgylisawpd 4170
        |||
Qy      406 --PDAAHVAVNLSRYEG 419
        |||
Db      418 slpdlstsvfnqlvirlg 433

```

XX	AAW92406	standard; Protein; 1255 AA.
XX	AAW92406;	
XX	21-APR-1999	(first entry)
XX	Human HER-2/neu oncogene protein.	
XX	HER-2/neu; oncogene; immune response; T cell; B cell; immunisation;	
XX	malignancy; treatment; tumour.	
XX	Homo sapiens.	
XX	Key	Location/Qualifiers
XX	Region	676..1255
XX		/note="region which elicits immune response"
XX	US5869445-A.	
XX	09-FEB-1999.	
XX	01-APR-1996;	96US-0625101.
XX	01-APR-1996;	96US-0625101.
XX	17-MAR-1993;	93US-0033644.
XX	12-AUG-1993;	93US-0106112.
XX	31-MAR-1995;	95US-0414417.
XX	(UNIV) UNIV WASHINGTON.	
XX	Cheever MA, Disis ML;	
XX	WPI; 1999-152835/13.	
XX	N-PSDB; AAX01912.	
XX	Use of HER-2/neu polypeptides - for eliciting an immune response to	
XX	an HER-2/neu associated malignancy, particularly for treating or	
XX	preventing tumours	
XX	Claim 3; Column 31-38; 26pp; English.	
XX	This sequence represents the human HER-2/neu oncogene protein. A fragment	
XX	of this protein is used in a method for eliciting or enhancing an immune	
XX	response to HER-2/neu protein. The polypeptide can stimulate T cells and	
XX	B cells to produce an immune response to the HER-2/neu protein. The	
XX	method can be used for immunisation against a malignancy in which the	
XX	HER-2/neu oncogene is associated and in the treatment of an existing	
XX	tumour, or to prevent tumour occurrence or reoccurrence.	
XX	Sequence 1255 AA;	
XX	Query Match	82.1%; Score 1878; DB 20; Length 1255;
XX	Best Local Similarity	83.0%; Pred. No. 4e-143;
XX	Matches 362; Conservative	9; Mismatches 45; Indels 20; Gaps 5

PR 02-OCT-1998; 98US-0165192.
 XX (UYMC-) UNIV MCMMASTER.
 PA
 XX Muller WJ, Siegel PM;
 PI
 XX WPI: 2000-303768/26.
 DR N-PSDB: AAA14812.
 XX
 PT Nucleic acid encoding an erbB 2 receptor protein designated SPLICE
 PT erbB-2, inhibitors of the protein are useful for treatment of cancer
 PS
 PS Claim 3; Fig 2; 60pp; English.
 CC
 CC The present sequence represents a SPLICE erbB-2 receptor protein. The
 CC protein has an in-frame deletion of 16 amino acids, 2 of which are
 CC conserved cysteine residues, compared to the unspliced protein. The
 CC erbB-2 polynucleotide is used to construct probes for detecting
 CC disorders of cell transformation such as cancer. Antibodies to the
 CC protein may be used to detect SPLICE erbB-2 in a sample. Agents
 CC (e.g. antisense oligonucleotides) which inhibit the expression of
 CC SPLICE erbB-2 are useful for reducing tumor cell proliferation and
 CC treating cancer. Substances which stimulate SPLICE erbB-2 are useful
 CC for treating conditions of tissue occurs, such as arthropathy, bone
 CC resorption, inflammatory diseases, degenerative disorders of the
 CC central nervous system and wound healing.
 CC
 CC
 SQ Sequence 1255 AA;

Query Match 82.1%; Score 1878; DB 21; Length 1255;
 Best Local Similarity 83.0%; Pred. No. 4e-143;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELAALCRLGALLLALPPGASTVCTGDMKRLPASPETHLDMRLHYGCGVQVQNL 60
 DB 1 melaalcrvgllallppgaastvctgtdmklrpsapethldmrlhygcvqvvgnl 60
 QY 61 EITLYPTNASTSLFDIDIQVQGVLIANQVROVPLQRLRYRGTOLEFDNTALAVLDNG 120
 DB 61 eltylptnaslsfldidqevgvvliahnqvrvpqlrlyrgtqlfednyalavldng 120
 QY 121 DPLNNTPVTGASPGGLRLOLRSLTEIKGVLIOBNOLCYODITLKKDIFHKNNOLA 180
 DB 121 dplnntpvtgaspgglrlolrslteikgvlionbolyoditlkkdifhknnoia 180
 QY 181 LTLIDTNRSRACHPCSPMKSGRCMGESESDCOSLTRTVACAGACARCKGPLPTDCHEQC 240
 DB 181 ltlidtnrsrachpcspmksgrcmgesesdcosltrtvacagacarkgplptdccheqc 240
 QY 241 AAGCGPKKSDCLACLFHNSGICELHCPALVYNTVFESMPNPGRYTCASCVTACP 300
 DB 241 aagcgpkkdsdclacLFHNSGICELHCPALVYNTVFESMPNPGRYTCASCVTACP 300
 QY 301 YVYLSDVQSCTLVCPILHNOEYTAEDGTORCEKSKPCAR-----GTHSLRPAAVPVP 355
 DB 301 yvylsdvqscTLVCPILHNOEYTAEDGTORCEKSKPCAR-----GTHSLRPAAVPVP 355
 QY 356 LRMQGC--PAHFVLSFLRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRPD 405
 DB 356 lrmqgc--PAHFVLSFLRPSMDLVSAFYSLPLAPLSPTSVPI-----SPVSVGRPD 405
 QY 361 lqefagckkifgsiaflpfesfdgpaantc--apilqpeqlqfclleeltlyllylsawpd 417
 DB 361 lqefagckkifgsiaflpfesfdgpaantc--apilqpeqlqfclleeltlyllylsawpd 417
 QY 406 --PDAHVAVNLSRYEG 419
 DB 418 slpdlsvlfnqlqvtrg 433

RESULT 13
 ID AAY92620
 XX AAY92620 standard; Protein; 1255 AA.
 AC AAY92620;

XX 10-AUG-2000 (first entry)
 DT
 XX
 DE Human heregulin 2 (Her2).
 XX
 KW Heregulin 2; Her2; vaccination; cytotoxic T-lymphocyte immunity;
 KW self-protein; cancer; breast cancer; prostate cancer;
 KW cell-associated peptide antigen; foreign epitope.
 XX
 OS Homo sapiens.
 XX
 FH Key
 FH Domain
 FT
 FT Location/Qualifiers
 FT 1..173
 FT /label= "N-terminal
 FT /note= "mature polypeptide"
 FT 5..25
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 59..73
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 103..117
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 149..163
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 174..323
 FT /label= "Cysteine-rich_domain
 FT 210..224
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 250..264
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 324..483
 FT /label= "Ligand_binding_domain
 FT 325..339
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 369..383
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 465..479
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 484..623
 FT /label= "Cysteine-rich_domain
 FT 579..593
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 624..654
 FT /label= "Transmembrane_domain
 FT 632..652
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 653..667
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 655..1010
 FT /label= "Tyrosine_kinase_domain
 FT 661..675
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 695..709
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 710..730
 FT /label= "insertion_region
 FT /note= "suitable for foreign epitope insertion"
 FT 1011..1235
 FT /label= "C-terminal_domain
 XX

PN WO200020027-A2.
XX
PD 13-APR-2000.
XX
PF 05-OCT-1999; 99WO-DK00525.
XX
PR 05-OCT-1998; 98DK-0001261.
PR 20-OCT-1998; 98US-0105011.
XX
PA (MEBI-) M & E BIOTECH AS.
XX
PI Steinaa L, Mouritsen S, Nielsen KG, Haaning J, Leach D, Dalum I,
PI Gautam A, Birk P, Karlsson G;
XX
DR WPI: 2000-349917/30.
XX N-PSDB; AAA09455.
XX
PT Inducing immune responses to weakly immunogenic, tumor associated
PT peptide antigens for the treatment of breast and prostate cancer
XX
PS Claim 62; Page 193-198; 220pp; English.
XX
CC This is the human heregulin 2 (Her2) sequence. Immunogenic analogues of
CC Her2 can be used in the claimed method as an autovaccine to induce a CTL
CC response. Subdominant CTL epitopes, antibody binding regions and
CC cysteine residues involved in disulfide bonds are preserved in the
CC immunogenized forms. Regions suitable for the insertion of foreign T
CC helper epitopes were identified (see features table). The method
CC is used for inducing immune responses against weakly immunogenic
CC cell-associated peptide antigens (PA) such as those associated with
CC cancers (self-proteins), e.g. human prostate specific membrane antigen
CC (PSM), heregulin 2 (Her2) and/or fibroblast growth factor 8b (FGF8b).
CC The method comprises effecting simultaneous presentation by antigen
CC producing cells (APCs) of the animals immune system of: (1) at least 1
CC CTL (cytotoxic T-lymphocyte) group derived from the PA and/or at least 1
CC B-cell group derived from the cell-associated PA; and (2) at least 1
CC first T helper cell group which is foreign to the animal. Analogues of
CC human PSM, human Her2 and human/murine FGF8b comprising a substantial
CC part of all known and predicted CTL and B-cell epitopes of the respective
CC PA and including at least one foreign T helper epitope are also claimed.
CC The method is used to treat prostate, prostate/breast or breast cancer.
CC When the PA is human PSM, FGF8b and Her2, respectively.
XX
SQ Sequence 1255 AA;
Query Match 82.1%; Score 1878; DB 21; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
OY 1 METAALCRWGLLALLPRAASTQVCTGDMKRLPASPETHLDMRLHYGCGQVVGNL 60
DB 1 metaalcrrgllallppraastqvcctgctdmkrlrpsapethldmrlhygscqvvgnl 60
OY 61 ELTYLPNTNASLSEFLDIOGVGVLLAHNOVROPLORLRIYRGTOLEFDNVALAVLNDG 120
DB 61 eltylpntnaslsflgdigvgvyliahnqvrvpqlrlyrgtqlfednyalavldng 120
OY 121 DPLNNTPTVTGASPGRLRELQLRSLTEILKGVLIORNPOLCYODITLKKDFHRNNOLA 180
DB 121 dplnntptvtgaspgrlrelqlrslteilkgvliornpolecyoditlkkdfhrnnola 180
OY 121 dplnntptvtgaspgrlrelqlrslteilkgvliornpolecyoditlkkdfhrnnola 180
DB 181 LTIIDNRBRACHPCSPMKGSRMGESSDQSLRTVCAGCAGCKGRLPTDCHECC 240
OY 181 ltiidnrbrachpcspmkgsrmgessdqslrtvcagcagckgrrlptdchecc 240
DB 181 ltiidnrbrachpcspmkgsrmgessdqslrtvcagcagckgrrlptdchecc 240
OY 241 AAGCTGPKRSDCLACLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCTYACP 300
DB 241 aagctgpkrsdclacLFHNSGICELHCPALVYNTDFESMPNPGRYTFGASCTYACP 300
OY 301 YNTLSTDVSCSTLVCPLAHNOEVTAEDEGTORCEKSKPCAR----GTHSLPRPAAVPV 355
DB 301 yntlstdvscstlvcplahnoevtaedegtorcekscpcar----gthslprpaavpv 355
DB 301 yntlstdvscstlvcplahnoevtaedegtorcekscpcarvcyglgmhlrevravtsan 360

OY 356 LRMPG--PAHPVLSFLRPSMDVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 361 lqefagckkifsglsflrpsfmdvsaFYSLPLAPLSPTSVPI-----SPVSGRGPD 417
OY 406 --PDAHVAVNLSTRYEG 419
DB 418 slpdsfvgfnglqvtrg 433
RESULT 14
AAB85458
ID AAB85458 standard; Protein; 1255 AA.
XX
AC AAB85458;
XX
DT 25-SEP-2001 (first entry)
XX
DE Human HER-2/neu protein.
XX
KW Antigen-presenting cell; immunogenic; immune response; HER-2/neu;
KW oncogene; cancer; cytostatic; vaccine; p185; c-erbB2.
XX
OS Homo sapiens.
XX
PF WO200153463-A2.
XX
PD 26-JUL-2001.
XX
PI 19-JAN-2001; 2001WO-US01850.
XX
PR 21-JAN-2000; 2000US-0177545.
XX
PA (CORI-) CORIXA CORP.
XX
PI Cheever MA, Hand-Zimmermann S;
XX
DR WPI: 2001-476112/51.
XX N-PSDB; AAH23392.
PT New antigen-presenting cells, useful as vaccines for eliciting or
PT enhancing an immune response to HER-2/neu protein, particularly useful
PT for treating or preventing cancer, e.g. breast cancer
XX
PS Claim 2; Page 41-46; 49pp; English.
XX
CC The invention provides an isolated antigen-presenting cell, which
CC expresses at least an immunogenic portion of a polypeptide that produces
CC an immune response to HER-2/neu protein. The antigen-presenting cells are
CC useful as vaccines for eliciting or enhancing an immune response to
CC HER-2/neu protein, particularly in treating or preventing malignancies in
CC which the HER-2/neu oncogene is associated. Specifically, these are
CC useful for treating or preventing cancer, e.g. breast cancer, ovarian,
CC colon, lung or prostate cancers. The present sequence represents
CC the human HER-2/neu protein (also known as p185 or c-erbB2).
XX
SQ Sequence 1255 AA;
Query Match 82.1%; Score 1878; DB 22; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
OY 1 METAALCRWGLLALLPRAASTQVCTGDMKRLPASPETHLDMRLHYGCGQVVGNL 60
DB 1 metaalcrrgllallppraastqvcctgctdmkrlrpsapethldmrlhygscqvvgnl 60
OY 61 ELTYLPNTNASLSEFLDIOGVGVLLAHNOVROPLORLRIYRGTOLEFDNVALAVLNDG 120
DB 61 eltylpntnaslsflgdigvgvyliahnqvrvpqlrlyrgtqlfednyalavldng 120
OY 121 DPLNNTPTVTGASPGRLRELQLRSLTEILKGVLIORNPOLCYODITLKKDFHRNNOLA 180
DB 121 dplnntptvtgaspgrlrelqlrslteilkgvliornpolecyoditlkkdfhrnnola 180

DB 121 dplnltptvgaspqglrelqlrsteilkqgvliqnrpqlyqdtllwkdifhknqja 180
QY 181 LFLIDTNSRACHPCSPMKGRGCGESSEDCQSLTRFYCAGGCAKCKGRLPTDCHEQC 240
DB 181 lllldtnsrachpcspmkgrcgessedcqslltrvcaggarckypjldccheqc 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
DB 241 aagctgpkhsdclaclhfnhsigicelhcpalvtlyntdfesmpnpegryltfgascvtacp 300
QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCAR-----GTHSLPRPAVPVP 355
DB 301 ynylstdvgsctlvcpplhnqevtaedgtqrckskpcarvgylgmehlrevravtsan 360
QY 356 LRMQPG--PAHNVLSFLRSPMDLVSAFYSLPLAPISPTSVPT-----SPVSYGRGPD 405
DB 361 lqefagckkifgslafipdestidgdpasnt---apilqpeqlqvfeclteeltgylisawpd 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 slpdlsvfqnlgvtrg 433

RESULT 15
AAG88267
ID AAG88267 standard; Protein; 1255 AA.
XX
AC AAG88267;
XX
DT 11-SEP-2001 (first entry)
XX
DE HER2/neu amino acid sequence.
XX
KW Human; HER2/neu; epitope; human leukocyte antigen; HLA; T cell;
KW Immune response; vaccine; tumour; cancer; cytostatic; immunostimulant;
KW tumour-associated antigen; T lymphocyte; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
PN WO200141787-A1.
XX
PD 14-JUN-2001.
XX
PF 11-DEC-2000; 2000WO-US33591.
XX
PR 10-DEC-1999; 99US-0458299.
XX
PA (EPIM-) EPIMMUNE INC.
XX
PI Fikes J, Sette A, Sidney J, Southwood S, Chesnut R, Celis E;
PI Keogh E;
XX
DR WPI: 2001-374995/39.
XX
PT An isolated prepared HER2/neu epitope useful in a vaccine for inducing
PT cellular immune responses for the prevention and treatment of cancer -
XX
PS Disclosure: Page 15; 1999p: English.

XX The present invention describes isolated prepared HER2/neu epitopes (I).
CC Also described are: (1) a clonal cytotoxic T lymphocyte (CTL) that is
CC culture in vitro and binds to a complex of an epitope (I), bound to a
CC human leukocyte antigen (HLA) molecule; (2) a peptide (II) comprising (I)
CC and a second epitope and the peptide is less than 50 contiguous amino
CC acids that have 100% identity with a native peptide sequence of HER2/neu;
CC (3) a vaccine composition (III) comprising (II) and a pharmaceutical
CC excipient; (4) an isolated nucleic acid encoding a peptide comprising
CC (I); and (5) an isolated nucleic acid encoding (II). (I) has cytostatic
CC and immunostimulant activities, and can be used in vaccines. (I), (II)
CC and (III) are useful for inducing cellular immune responses for the
CC prevention and treatment of cancer. (I) and (II) are useful for
CC monitoring or evaluating an immune response to a tumour-associated
CC antigen when incubated with a T lymphocyte sample form a patient and

CC detecting the presence of bound T lymphocyte to (I) or (II). Epitope
CC based vaccines mean that immunosuppressive epitopes that may be present
CC in whole antigens may be avoided. Selected pathological side effects caused by
CC enhance immunogenicity. The possible pathological side effects caused by
CC infectious agents or whole protein antigen is eliminated. The vaccine
CC provides the ability to direct and focus an immune response to multiple
CC selected antigens from the same pathogen. Epitope-based anti-tumour
CC vaccines provides the opportunity to combine epitopes derived from
CC multiple tumour-associated molecules addressing the problem of tumour-
CC tumour variability and reducing the likelihood of tumour escape due to
CC antigen loss. AAG88266 to AAG89121 represent amino acid sequences used in
CC the exemplification of the present invention.

SO Sequence 1255 AA;

Query Match 82.1%; Score 1878; DB 22; Length 1255;
Best Local Similarity 83.0%; Pred. No. 4e-143;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALCRWGLLALLPGCASTQVCTGDMKRLPASPETHLDMRLHYGCGYVQGNL 60
DB 1 melalcrwglallallpgcaastyctgtdmklrllpaspethldmrlhygcyvqvgnl 60
QY 61 ELTYLPTNASLSFLDIOEVQGVLIANQVQVPLQRLRIYRGTOLPEDNYALAVLNG 120
DB 61 eltylptnaslsfldioevqgvyliahnqvqvplqrlrliryrgtqlpednyalavldng 120
QY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQNRPNOLCYODTILMKDIFHRKNOLA 180
DB 121 dplnntptvtgaspgglrelqrlslteilkqgvliqnrpnolcyodtllmkdifhknqja 180
QY 181 LFLIDTNSRACHPCSPMKGRGCGESSEDCQSLTRFYCAGGCAKCKGRLPTDCHEQC 240
DB 181 lllldtnsrachpcspmkgrcgessedcqslltrvcaggarckypjldccheqc 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
DB 241 aagctgpkhsdclaclhfnhsigicelhcpalvtlyntdfesmpnpegryltfgascvtacp 300
QY 301 YNYLSTDVGSCTLVCPPLHNQEVTAEDGTQRCCKSKPCAR-----GTHSLPRPAVPVP 355
DB 301 ynylstdvgsctlvcpplhnqevtaedgtqrckskpcarvgylgmehlrevravtsan 360
QY 356 LRMQPG--PAHNVLSFLRSPMDLVSAFYSLPLAPISPTSVPT-----SPVSYGRGPD 405
DB 361 lqefagckkifgslafipdestidgdpasnt---apilqpeqlqvfeclteeltgylisawpd 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 slpdlsvfqnlgvtrg 433

Search completed: April 11, 2002, 09:29:11
Job time: 53 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:30:07 ; Search time 25.01 Seconds
(without alignments)
1276.175 Million cell updates/sec

Title: US-09-234-208B-2
Perfect score: 2287
Sequence: 1 METALCRMGLLALTPGA.....VGRGPPDAHVAVNLSPREG 419

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	1255	1 A24571	protein-tyrosine k
2	1608.5	70.3	1260	1 TYRATNU	protein-tyrosine k
3	1571	68.7	1254	2 148161	p-185 precursor -
4	796.5	34.8	1223	1 TVCHLV	epidermal growth f
5	794	34.7	527	2 A42032	epidermal growth f
6	793	34.7	1210	1 GQHUE	epidermal growth f
7	789	34.5	1210	2 A53183	epidermal growth f
8	781.5	34.2	644	2 A36325	epidermal growth f
9	775	32.9	1308	2 A47253	epidermal growth f
10	735.5	32.2	1342	2 A36223	epidermal growth f
11	692	30.3	1339	2 JC4387	kinase-related tra
12	681.5	29.8	1166	1 S06142	epidermal growth f
13	575.5	25.2	843	2 A27131	epidermal growth f
14	419	18.3	1374	2 S70712	epidermal growth f
15	416	18.2	1369	2 S70713	protein-tyrosine k
16	415	18.1	1330	1 GQFE	epidermal growth f
17	363.5	15.9	366	2 D45558	epidermal growth f
18	363.5	15.9	1717	1 A45558	epidermal growth f
19	342.5	15.0	1363	2 T43220	insulin-like growt
20	331	14.5	333	2 B45558	epidermal growth f
21	331	14.5	342	2 C45558	epidermal growth f
22	294	12.9	2101	2 S57245	insulin receptor f
23	294	12.9	2148	1 A56081	insulin receptor f
24	291	12.7	1477	2 T18534	protein-tyrosine k
25	278	12.2	1300	2 A36502	insulin receptor-r
26	272	11.9	540	2 B47417	insulin receptor-r
27	269.5	11.8	1607	2 T43212	insulin-like growt
28	263.5	11.5	1382	1 INHUR	insulin receptor p
29	263.5	11.5	1383	2 A36080	insulin receptor p

30	261	11.4	1372	2 A34157	insulin receptor p
31	258	11.3	1390	2 T30346	insulin receptor -
32	257.5	11.3	1367	1 IGHUPL	insulin-like growt
33	252.5	11.0	1371	2 A33837	insulin-like growt
34	249	10.9	1268	2 B36502	insulin receptor-r
35	235	10.3	329	2 A48805	tyrosine kinase re
36	226.5	9.9	183	2 JH0803	insulin receptor h
37	203	8.9	1846	2 T42047	insulin receptor h
38	160.5	7.0	1299	2 T43251	furin (EC 3.4.21.7
39	144	6.3	1548	2 S34583	serine proteinase
40	142	6.2	1959	1 AGRT	agrin - rat
41	136.5	6.0	1111	2 T26972	hypothetical prote
42	133.5	5.8	915	1 A48225	subtilisin-like pr
43	132	5.8	1574	2 T13954	MEGF6 protein - ra
44	131.5	5.7	915	2 B48225	proble proteotet
45	131.5	5.7	915	2 JC6148	subtilisin-like pr

ALIGNMENTS

RESULT 1
A24571
protein-tyrosine kinase (EC 2.7.1.112) erbB2 precursor - human
N:Alternate names: c-erb-B-2 protein precursor; Kinase-related transforming protein e
C:Species: Homo sapiens (man)
C:Date: 25-Oct-1987 #sequence_revision 06-Dec-1996 #text_change 11-Jun-1999
C:Accession: A24571; A25491; A44188; B44188; I59509; I57622
R:Yamamoto, T.; Ikawa, S.; Akiyama, T.; Semba, K.; Nomura, N.; Miyajima, N.; Saito, T
Nature 319, 230-234, 1986
A:Title: Similarity of protein encoded by the human c-erb-B-2 gene to epidermal growth
A:Reference number: A24571; MUID:86118663
A:Accession: A24571
A:Molecule type: mRNA
A:Residues: 1-1255 <YAM>
A:Cross-references: GB:X03363; NID:931197; PIDN:CAA27060.1; PID:931198
R:Semba, K.; Kamata, N.; Toyoshima, K.; Yamamoto, T.
Proc. Natl. Acad. Sci. U.S.A. 82, 6497-6501, 1985
A:Title: A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epid
A:Reference number: A25491; MUID:86016729
A:Accession: A25491
A:Molecule type: DNA
A:Residues: 737-1031 <SEM>
A:Cross-references: GB:M1767; NID:g182163; PIDN:AAA35808.1; PID:g553282
R:Cousens, L.; Yang-Feng, T.L.; Liao, Y.C.; Chen, E.; Gray, A.; McGrath, J.; Seeburg
Science 230, 1132-1139, 1985
A:Title: Tyrosine kinase receptor with extensive homology to EGF receptor shares chro
A:Reference number: A44188; MUID:86070181
A:Accession: A44188
A:Molecule type: DNA
A:Residues: 740-910 <COU1>
A:Cross-references: GB:M12036; NID:g183988; PIDN:AAA35978.1; PID:g183989
A:Accession: B44188
A:Molecule type: mRNA
A:Residues: 1-517; 'RALI', 522, 'S', 524-654, 'V', 656-1169, 'A', 1171-1255 <COU2>
A:Cross-references: GB:M11730; NID:g183986
R:King, C.R.; Kraus, M.H.; Aaronson, S.A.
Science 229, 974-976, 1985
A:Title: Amplification of a novel v-erbB-related gene in a human mammary carcinoma.
A:Reference number: I59509; MUID:85272597
A:Accession: I59509
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 837-909 <REX>
A:Cross-references: GB:I29395; NID:g459807; PIDN:AAA35809.1; PID:g459808
R:Tal, M.; King, C.R.; Kraus, M.H.; Ullrich, A.; Schlessinger, J.; Givol, D.
Mol. Cell. Biol. 7, 2597-2601, 1987
A:Title: Human HER2 (neu) promoter: evidence for multiple mechanisms for transcriptio
A:Reference number: I57622; MUID:87286898
A:Accession: I57622
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-191 <TAU>

A:Cross-references: GB:M16792; NID:9183983; PIDN:AA58637.1; PID:9553332
 C:Comment: Amplification and overexpression of this erbB-related gene occurs in about 30
 C:Genetics: GDB:ERBB2; NGU; HER-2
 A:Gene: GDB:ERBB2; NGU; HER-2
 A:Cross-references: GDB:120613; OMIM:164870
 A:Map position: 17q21.1-17q21.1
 A:Introns: 25/1; 75/3; 147/1; 883/3
 A>Note: the list of introns is incomplete
 C:Function: catalyzes the phosphorylation of a peptidyl tyrosine residue by ATP
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phosphatase
 F:1-21/Domain: signal sequence #status predicted <SIG>
 F:22-125/Domain: protein-tyrosine kinase erbB2 #status predicted <MAT>
 F:22-653/Domain: extracellular #status predicted <EXT>
 F:70-304/Domain: EGF receptor extracellular domain repeat <EE1>
 F:395-605/Domain: EGF receptor extracellular domain repeat <EE2>
 F:654-675/Domain: transmembrane #status predicted <TM>
 F:676-1255/Domain: intracellular #status predicted <INT>
 F:718-983/Domain: protein kinase homology <KIN>
 F:726-734/Region: protein kinase ATP-binding motif
 F:686-124,187,253,530,571,629/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:686/Binding site: phosphate (Thr) (covalent) (by protein kinase C) #status predicted
 F:753/Active site: Lys #status predicted
 F:1139,1221,1222,1248/Binding site: phosphate (Tyr) (covalent) (by autophosphorylation)

Query Match 82.1%; Score 1878; DB 1; Length 1255;
 Best Local Similarity 83.0%; Pred. No. 4,2e-121;
 Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

OY 1 MELAALCRNGILLALLPFGAASVQVCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60
 DB 1 MELAALCRNGILLALLPFGAASVQVCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60
 OY 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
 DB 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
 OY 121 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 180
 DB 121 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 180
 OY 121 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 180
 DB 121 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 180
 OY 181 LTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 240
 DB 181 LTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 240
 OY 181 LTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 240
 DB 181 LTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 240
 OY 241 AAGCTGPKHSDDLACHLHNSGICELHCPALVYNTDFESMPNDEGRYTFGASCVTAC 300
 DB 241 AAGCTGPKHSDDLACHLHNSGICELHCPALVYNTDFESMPNDEGRYTFGASCVTAC 300
 OY 301 YNTLSTDVGSCTIVCPHLNDEVTAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
 DB 301 YNTLSTDVGSCTIVCPHLNDEVTAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
 OY 301 YNTLSTDVGSCTIVCPHLNDEVTAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
 DB 301 YNTLSTDVGSCTIVCPHLNDEVTAEDGTQRCCKSPCAR-----GTHSLPRAVP 355
 OY 356 LRMQPP--PAHPLVSLFRPSMDLVSAFYSPLAPLSPTSVPI-----SPVSGR 405
 DB 356 LRMQPP--PAHPLVSLFRPSMDLVSAFYSPLAPLSPTSVPI-----SPVSGR 405
 OY 361 IQEFACCKKIFGSLAFPLPESFDDPASNT---APLQPEQLQVLEETITGYLISAMP 417
 DB 361 IQEFACCKKIFGSLAFPLPESFDDPASNT---APLQPEQLQVLEETITGYLISAMP 417
 OY 406 --PDAAHVAVNLRYREG 419
 DB 418 SLPDLSPFONLQVIRG 433

RESULT 2
 TYRTNU
 protein-tyrosine kinase (BC 2.7.1.112) neu precursor - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 31-Dec-1988 #sequence_revision 31-Dec-1988 #text_change 11-Jun-1999
 C:Accession: A24562; A61204
 R:Bargmann, C.I.; Hung, M.C.; Weinberg, R.A.
 Nature 319, 226-230, 1986

A:Title: The neu oncogene encodes an epidermal growth factor receptor-related protein
 A:Reference number: A24562; MUID:86118662
 A:Accession: A24562
 A:Molecule type: mRNA
 A:Residues: 1-1260 <EMBL>
 A:Cross-references: EMBL:X03362; NID:956745; PIDN:CA27059.1; PID:956746
 R:Masui, T.; Mann, A.M.; Macatee, T.L.; Garland, E.M.; Okamura, T.; Smith, R.A.; Cole
 Carcinogenesis 12, 1975-1978, 1991
 A:Title: Direct DNA sequencing of the rat neu oncogene transmembrane domain reveals a
 2-thiazolylformamide or N-methyl-N-nitrosourea.
 A:Reference number: A61204; MUID:92035293
 A:Accession: A61204
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 637-663, 'V', 665-702 <MAS>
 A>Note: authors translated the codon GCA for residue 25 as Val
 C:Genetics:
 A:Gene: neu
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
 F:1-19/Domain: signal sequence #status predicted <SIG>
 F:20-1260/Domain: protein-tyrosine kinase neu #status predicted <MAT>
 F:658-680/Domain: transmembrane #status predicted <TM>
 F:723-988/Domain: protein kinase homology <KIN>
 F:731-739/Region: protein kinase ATP-binding motif
 F:71,191,263,535,576,634/Binding site: carbohydrate (Asn) (covalent) #status predicted
 F:691/Binding site: phosphate (Thr) (covalent) #status predicted
 F:758/Active site: Lys #status predicted
 F:882,1227,1253/Binding site: phosphate (Tyr) (covalent) #status predicted

Query Match 70.3%; Score 1608.5; DB 1; Length 1260;
 Best Local Similarity 85.0%; Pred. No. 1.3e-102;
 Matches 294; Conservative 15; Mismatches 36; Indels 1; Gaps 1;

OY 1 MELAALCRNGILLALLPFGAASVQVCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 60
 DB 4 MELAALCRNGILLALLPFGAASVQVCTGDMKRLPASPETHLDMRLHYOGCQVQGNL 63
 OY 61 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 120
 DB 64 ELTYLPNTASLFLDIOEVOGVYLLAHNOVROVPLQRLRYRGTOLEFEDNALAVLNDG 123
 OY 121 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 179
 DB 124 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 183
 OY 121 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 179
 DB 124 DPLNNTPTVYGASPGGLREQLRLSLTEILKGVLIQNRPOLCYQDTILMKDIFKKNQLA 183
 OY 180 ALTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 239
 DB 184 ALTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 243
 OY 180 ALTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 239
 DB 184 ALTLIDITNRSRACHPCSPMKGRCWGESSEDCQSLRTVCAGGCAKCKGRLPTDCCHQ 243
 OY 240 CAAGCTGPKHSDDLACHLHNSGICELHCPALVYNTDFESMPNDEGRYTFGASCVTAC 299
 DB 244 CAAGCTGPKHSDDLACHLHNSGICELHCPALVYNTDFESMPNDEGRYTFGASCVTAC 303
 OY 300 PYNYLSTDVGSCTIVCPHLNDEVTAEDGTQRCCKSPCARSTHSL 345
 DB 304 PYNYLSTDVGSCTIVCPHLNDEVTAEDGTQRCCKSPCARSTHSL 349

RESULT 3
 I48161
 p-185 precursor - golden hamster
 C:Species: Mesocricetus auratus (golden hamster)
 C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 18-Jun-1999
 C:Accession: I48161
 R:Nakamura, T.; Ushijima, T.; Ishizaka, Y.; Nagao, M.; Arai, M.; Yamazaki, Y.; Ishika
 Gene 140, 251-255, 1994
 A:Title: Cloning and activation of the Syrian hamster neu proto-oncogene.
 A:Reference number: I48161; MUID:94193007
 A:Accession: I48161
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA

Db 80 EHNRLDTFLTKTIOEVAGYVLIANMADVIPLEMLQITIRGNVLYDNSFALAVLSNYH-NMK 138
 Oy 126 TTPVNTASPGJBELOLRSLTEILKGCYLQIRNPOLCYODTLIMKRIFFKNNOLATLID 185
 Db 139 TO-----GLLELPKRLSELINGVKISNNPKLCNMDTVLMDIIDTSRK-PLTVAD 189
 Oy 186 -TNRSHACHPCSPCKSGKSGKMGESSEDCOSLRTVYAGCA-RCKGPLEPTDCHEQCAAG 243
 Db 190 FASNLSSCKPKCHNCTEDHOMGAGEQNCOTLRFVICAQCSGRCRKVPSDCCHNCAG 249
 Oy 244 CTGPKRSDICLAFHNSGICELHCPALVTYNTDFESAPNPEGRTFGASCYTACPYNY 303
 Db 250 CTGPRSDCLACKRFDDATCKTCEPLVLYNTTYQMVPNEGKYSFGATCVRECPHNY 309
 Oy 304 LSTDVSGCTLYVCLHNOEYTAEDGTORCEKSGPCAR 340
 Db 310 VYTDHGSVRSCTNTDTEY-EEGVRKCKKCGDLCSK 345

RESULT 6 GOMUE

epidermal growth factor receptor precursor - human
 N:Contains: protein-tyrosine kinase (EC 2.7.1.112) erBb
 C:Species: Homo sapiens (man)
 C:Date: 15-Nov-1984 #sequence-revision 27-Nov-1985 #text-change 11-Jun-1999
 C:Accession: A00641; A25772; S30024; A38672; A06642; A43615; A23062; A05281; A60143; A33
 R:Ulrich, A.; Coussens, L.; Hayflick, J.S.; Dull, T.J.; Gray, A.; Tam, A.W.; Lee, J.; X
 tg, P.H.
 Nature 309, 418-425, 1984
 A:Title: Human epidermal growth factor receptor cDNA sequence and aberrant expression of
 A:Reference number: A00641; MUID:84219729
 A:Accession: A00641
 A:Molecule type: mRNA
 A:Residues: 1-1210
 A:Cross-References: EMBL:X00588; NID:g31113; PIDN:CAA25240.1; PID:g757924
 A:Note: the authors translated the codon AAG for residue 540 as Asn
 R:Shih, S.; Xu, Y.; Stratton, R.H.; Roe, B.A.; Merlino, G.T.; Pastan, I.
 Proc. Natl. Acad. Sci. U.S.A. 82, 4920-4924, 1985
 A:Title: Characterization and sequence of the promoter region of the human epidermal gro
 A:Reference number: A25772; MUID:85270438
 A:Accession: A25772
 A:Status: translation not shown
 A:Molecule type: DNA
 A:Residues: 1-29 <ISH>
 A:Cross-References: GB:M1234; NID:g181981; PIDN:AA52370.1; PID:g553272
 R:Halay, J.; Whittle, N.; Bennett, P.; Kinchington, D.; Ulrich, A.; Waterfield, M.
 Oncogene Res. 1, 375-396, 1987
 A:Title: The human EGF receptor gene: structure of the 110 kb locus and identification
 A:Reference number: S30024; MUID:88217333
 A:Accession: S30024
 A:Molecule type: DNA
 A:Residues: 1-29 <HA2>
 A:Cross-References: EMBL:X06370; NID:g31118; PIDN:CAA29668.1; PID:g31119
 R:Halay, J.D.; Waterfield, M.D.
 J. Biol. Chem. 266, 1746-1753, 1991
 A:Title: Contributory effects of de Novo transcription and premature transcript terminat
 A:Reference number: A38672; MUID:91107677
 A:Accession: A38672
 A:Molecule type: DNA
 A:Residues: 1-29 <HAL>
 A:Cross-References: GB:M8425; NID:g181977; PIDN:AA63171.1; PID:g553271
 A:Experimental source: carcinoma cell line A431-7
 R:Xu, Y.; Ishii, S.; Clark, A.J.L.; Sullivan, M.; Wilson, R.K.; Ma, D.P.; Roe, B.A.; Mer
 Nature 309, 806-810, 1984
 A:Title: Human epidermal growth factor receptor cDNA is homologous to a variety of RNAs
 A:Reference number: A00642; MUID:84245835
 A:Accession: A00642
 A:Molecule type: mRNA
 A:Residues: 'RCMWRRA', 150-187, 'KSVIOAV', 195, 'W', 197, 'A', 199-222, 'S', 224-304, 'RA', 307-321
 '798-799, 'ND', 802-811, 'R', 813-942 <XU>
 A:Experimental source: A431 human carcinoma cells, which have large numbers of EGF recep
 R:Lin, C.R.; Chen, W.S.; Krueger, W.; Stolarsky, L.S.; Weber, W.; Evans, R.M.; Verma, I.

Science 224, 843-848, 1984
 A:Title: Expression cloning of human EGF receptor complementary DNA: gene amplificati
 A:Reference number: A43615; MUID:84196372
 A:Accession: A43615
 A:Molecule type: mRNA
 A:Residues: 713-964 <LIN>
 A:Experimental source: epidermoid carcinoma cell line A431
 R:Simmen, F.A.; Gope, M.L.; Schultz, T.Z.; Wright, D.A.; Carpenter, G.; O'Malley, B.W.
 Biochem. Biophys. Res. Commun. 124, 125-132, 1984
 A:Reference number: A23062; MUID:85046483
 A:Accession: A23062
 A:Molecule type: mRNA
 A:Residues: 1028-1210 <SIM>
 R:Weber, W.; Gull, G.N.; Speiss, J.
 Science 224, 294-297, 1984
 A:Reference number: A05281; MUID:84172183
 A:Accession: A05281
 A:Molecule type: protein
 A:Residues: 25-30, 'S', 32-51, 454-467 <WEB>
 R:Russo, M.W.; Lukas, T.J.; Cohen, S.; Staros, J.V.
 J. Biol. Chem. 260, 5205-5208, 1985
 A:Title: Identification of residues in the nucleotide binding site of the epidermal g
 A:Reference number: A60143; MUID:85182650
 A:Accession: A60143
 A:Molecule type: protein
 A:Residues: 740-744, 'X', 746-747 <RUS>
 R:Roczowski, B.; Mosig, G.; Cohen, S.
 Nature 309, 270-273, 1984
 A:Title: ATP-stimulated interaction between epidermal growth factor receptor and supe
 A:Reference number: A38023; MUID:84191554
 A:Contents: annotation; receptor activity
 A:Note: the EGF receptor (and other tyrosine kinases) can nick double-stranded DNA
 R:Chen, W.S.; Lazar, C.S.; Lund, K.A.; Welsh, J.B.; Chang, C.P.; Walton, G.M.; Der, C
 Cell 59, 33-43, 1989
 A:Title: Functional independence of the epidermal growth factor receptor from a doma
 A:Reference number: A3331; MUID:90003233
 A:Contents: annotation; internalization signal
 C:Comment: Binding of EGF to the receptor leads to internalization of the EGF-recepto
 C:Genetics:
 A:Gene: GDB:EGFR
 A:Cross-References: GDB:120610; OMIM:131550
 A:Map position: 7p12.3-7p12.1
 C:Superfamily: epidermal growth factor receptor; protein kinase homology
 C:Keywords: ATP; autophosphorylation; duplication; glycoprotein; phosphoprotein; phos
 F:1-24/Domain: signal sequence #status predicted <SIG>
 F:25-1210/Product: EGF receptor #status predicted <MAT>
 F:25-645/Domain: extracellular #status predicted <EXT>
 F:75-300/Domain: EGF receptor extracellular domain repeat <EE1>
 F:390-600/Domain: EGF receptor extracellular domain repeat <EE2>
 F:646-668/Domain: transmembrane #status predicted <TM>
 F:669-1210/Domain: intracellular #status predicted <INT>
 F:710-975/Domain: protein kinase homology <KIN>
 F:718-726/Region: protein kinase ATP-binding motif
 F:999-1046/Region: coated-pit mediated internalization signal
 F:1047-1210/Region: inhibitory
 F:128-175, 352, 413, 444, 528, 603/Binding site: carbohydrate (Asn) (covalent) #status pre
 F:745/Active site: Lys #status experimental

Query Match 34.7%; Score 793; DB 1; Length 1210;
 Best Local Similarity 45.3%; Pred. No. 9.7e-47;
 Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;
 Oy 11 LLLALLPPAA--SYOVCTGDMKRLRSPETHDMLRHLVYOGGOVYOGNLELYLPN 68
 Db 14 LLAALCPASRALEEKRVCGTGNKTLQIGFEDHPLSLQRMNCEVVALNLEITTVGRN 73
 Oy 69 ASLSFLDIOEVGYVLIANOVROVPIORLIRVGTOLFEDNVALAVLDNGDPLNNTTP 128
 Db 74 YDLSFLKTIQEVAGYVLIANLNTVERIPLENLOITINGNMYEASVALAVLSND----- 126
 Oy 129 VYGASDGLRELOLSLEFIILKGVLIQIRNPOLCYODTLIMKRIFFKNNOLATLIDNR 188

Db 127 ---ANKTGKLELPMRNLQELHGAVFNSNPALCNVESIQMDIVSDPLNSMNDPQNH 183
Oy 189 SNACHPGSMCKGSRCKGESSSEDCOSLRTVACGCA-RCKGPLPTDCCHECAAGCTGP 247
Db 184 LGSCKCKPSCPCNGSCMGWGEENCOKLTKITCAQCCSGCRCKSPSDCHNCAAGCTGP 243
Oy 248 KHSDDLACHFHNSGICELHCPALVTYNDPESMPNPEGRTFGASCVTACPYVYLSND 307
Db 244 RSDPLCRKRFDEATCKDTCPLMLYNTTYQMDVNPBGKISFGATCVKCKPRNYVTD 303
Oy 308 VGSCTVPLNMQEVTAEADGTQRCCKSPCAR 340
Db 304 HGSVRCAGCADSEYK-EEGCVKCKCKGCPCK 335
RESULT 7
A53183
epidermal growth factor receptor precursor - mouse
C:Species: Mus musculus (house mouse)
C:Date: 06-Jan-1995 #sequence revision 06-Jan-1995 #text change 18-Jun-1999
C:Accession: A53183; A43818; S24942; A28941; S45325; I49643
R:Blutcke, N.C.; Phillips, H.K.; Qiu, T.H.; Copeland, N.G.; Earp, H.S.; Jenkins, N.A.;
Genes Dev. 6, 399-413, 1994
A:Title: The mouse waved-2 phenotype results from a point mutation in the EGF receptor
A:Reference number: A53183; MUID:94170986
A:Accession: A53183
A:Molecule type: mRNA
A:Residues: 1-1210 <LUE>
A:Cross-references: GB:U03425
R:Avivi, A.; Lax, I.; Ullrich, A.; Schlessinger, J.; Givol, D.; Morse, B.
Oncogene 6, 673-676, 1991
A:Title: Comparison of EGF receptor sequences as a guide to study the ligand binding site
A:Reference number: A43818; MUID:91232866
A:Accession: A43818
A:Molecule type: mRNA
A:Residues: 1-714 <AVI>
A:Cross-references: GB:X59698
R:Eislinger, D.P.; Serrero, G.
submitted to the EMBL Data Library, June 1992
A:Reference number: S24942
A:Accession: S24942
A:Molecule type: mRNA
A:Residues: 969-971, 'K', 973-1115, 'D' <EIS>
A:Cross-references: EMBL:Z12608
R:Heisermann, G.J.; Gill, G.N.
J. Biol. Chem. 263, 13152-13158, 1988
A:Title: Epidermal growth factor receptor threonine and serine residues phosphorylated
A:Reference number: A28941; MUID:88330814
A:Accession: A28941
A:Molecule type: protein
A:Residues: 689-694, 'X', 696-704, 'L', 706-707, 989-992, 'XX', 995-996, 'X', 998-1000, 1002-1009,
R:Hibbs, M.L.; Dunn, A.R.; Alexander, W.S.
submitted to the EMBL Data Library, April 1994
A:Description: The complete cDNA sequence of the Mouse Epidermal Growth Receptor
A:Reference number: S45325
A:Accession: S45325
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-971, 'K', 973-1210 <VER>
A:Cross-references: EMBL:X78987; NID:G488830; PIDN:CAA5587.1; PID:G488831
R:Patla, B.C.; Das, S.K.; Andrews, G.K.; Dey, S.K.
Proc. Natl. Acad. Sci. U.S.A. 90, 55-59, 1993
A:Title: Expression of the epidermal growth factor receptor gene is regulated in mouse
A:Reference number: I49643; MUID:93126380
A:Accession: I49643
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 12-20, 22-132 <RES>
A:Cross-references: GB:I06864; NID:G193001; PIDN:AAA53029.1; PID:G567201
C:Genetics:
A:Gene: EGFR
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: Arp; growth factor receptor; kinase-related transforming protein; phosphoprotein

F:124/Domain: signal sequence #status predicted <SIG>
F:648-670/Domain: transmembrane #status predicted <TM>
F:712-977/Domain: protein kinase homology <KIN>
F:720-728/Region: protein kinase Arp-binding motif
F:680,695/Binding site: phosphate (Thr) (covalent) #status experimental
F:697,1070,1071/Binding site: phosphate (Ser) (covalent) #status experimental
F:993/Binding site: (or 997) phosphate (Ser) (covalent) #status experimental
F:1028/Binding site: (or 1030 or 1032) phosphate (Ser) (covalent) #status experimental
F:1197/Binding site: phosphate (Tyr) (covalent) #status experimental
Query Match 34.5%; Score 789; DB 2; Length 1210;
Best Local Similarity 46.3%; Pred. No. 1,86-46;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;
Oy 11 LIALLPFGAA-STQVCTGDMKRLPASPETHLDMRLHYQGVQVGNLELTPTN 68
Db 14 LITFALCAAGGALAEKKVCGQTSNRLTQGTEDHPLSIQRMVNNCEVYALNLEITYQRN 73
Oy 69 ASLSPLODIOEVQGVYLIANHOVQVPLQRLRYNGTQLFEDNVALAVLDNGDPLNNTTP 128
Db 74 YDLSPLKTIQEVAGVLIATLMTVERIPLENQIIRGNALVENTYALATLSN----- 124
Oy 129 VTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODITLWKDI---FHKNNOLATLI 184
Db 125 -YGTNRGTGLRELPMRNLOEILIGAVRFNNPILCMQDITQMRDIYQNVFNSMSMDL--- 180
Oy 185 DTRNSRACHPCSPMCKGSRCKGESSSEDCOSLRTVACGCA-RCKGPLPTDCCHEQCAAG 243
Db 181 -QSHPSCKPKDPPSCPNCSMGWGEENCOKLTKITCAQCCSHRCKGRSPSDCHNCAAG 239
Oy 244 CTGPRHSCCLACHFNHSGICELHCPALVTYNDPESMPNPEGRTFGASCVTACPYVY 303
Db 240 CTGPRSDCLVCKQKFDQDATCKDTCPLMLYNTTYQMDVNPBGKISFGATCVKCKPRNY 299
Oy 304 LSTDVGSCTVPLNMQEVTAEADGTQRCCKSPCAR 340
Db 300 VTDHGSVCVRACGPDYEV-EEGDKRCKCKDGPCK 335
RESULT 8
A36325
epidermal growth factor receptor - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 25-Jan-1991 #sequence revision 25-Jan-1991 #text change 10-Oct-1997
C:Accession: A36325
R:Petich, L.A.; Harris, J.; Raymond, W.W.; Blasband, A.; Lee, D.C.; Earp, H.S.
Mol. Cell. Biol. 10, 2973-2982, 1990
A:Title: A truncated, secreted form of the epidermal growth factor receptor is encoded
A:Reference number: A36325; MUID:90258888
A:Accession: A36325
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-644 <PEP>
A:Cross-references: GB:X37394
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: alternative splicing; Arp; growth factor receptor
Query Match 34.2%; Score 781.5; DB 2; Length 644;
Best Local Similarity 46.0%; Pred. No. 3,1e-46;
Matches 157; Conservative 44; Mismatches 117; Indels 23; Gaps 7;
Oy 3 LIALCRMGLLALLPFGA-STQVCTGDMKRLPASPETHLDMRLHYQGVQVGNLE 61
Db 15 LIALCAAG-----GALAEKKVCGQTSNRLTQGTEDHPLSIQRMVNNCEVYALNLE 66
Oy 62 LTIPLTNASLSFLDIOEVQGVYLIANHOVQVPLQRLRYNGTQLFEDNVALAVLDNGD 121
Db 67 ITTVQARNYDLSFLKTIQEVAGVLIATLMTVERIPLENQIIRGNALVENTYALAVLSN-- 124
Oy 122 PLNNTTPVTGASPGGLRELOLRSLTEILKGVLIQRNPOLCYODITLWKDI-FHKNNOLAT 181

Db 125 -----YGTNKTGLRELPMRLNLOELIGAVRFSNNPLTCNMETLQWEDIV-ODVFLSN 175
QY 182 TLIDNRRS-RACHPCSPMKGSRGCESESDCOSLTRVYACAGCA-RCKGPIPTCOCHQ 239
Db 176 MSMDVQRHLTGCPKCDPSPGNSGMRGECNCKLTKITCAQCCSRGSRPSDCCNQ 235
QY 240 CAAGCTGPKHSDCLALHNHSGICELHCPALVTYNTDFEPMNDEGRYTFGASCVTAC 299
Db 236 CAAGCTGPRSDPOLVCHRRDEATCKDPCPLMLYPTTYQMDVNPDEKYSFGACVKKC 295
QY 300 PNYLSTDVSGCTVCPPLHNOEYTAEDGQRCBCKSPCAR 340
Db 296 PNYVYTDHSGCVRACGPDYEV-EDDGVSKCKCKDCGPKRC 335

RESULT 9
A47253
epidermal growth factor receptor, HER4 - human
C:Species: Homo sapiens (man)
C:Date: 22-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 18-Jun-1999
C:Accession: A47253
R:Plowman, G.D.; Culloscu, J.M.; Whitney, G.S.; Green, J.M.; Carlton, G.W.; Foy, L.; Ne
Proc. Natl. Acad. Sci. U.S.A. 90, 1746-1750, 1993
A:Title: Ligand-specific activation of HER4/p180erbB4, a fourth member of the epidermal
A:Reference number: A47253; MUID:93189574
A:Accession: A47253
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: nucleic acid
A:Residues: 1-1308 <PILO>
A:Cross-references: GB:L07868; NID:9337359; PID:AA59446.1; PID:9337360
A:Note: sequence extracted from NCBI backbone (NCBIP:126842)
C:Superfamily: epidermal growth factor receptor; protein kinase homology
C:Keywords: ATP; growth factor receptor
F:716-981/Domain: protein kinase homology <KIN>
F:724-732/Region: protein kinase ATP-binding motif

Query Match 33.9%; Score 775; DB 2; Length 1308;
Best Local Similarity 45.1%; Pred. No. 1.8e-45;
Matches 157; Conservative 45; Mismatches 124; Indels 22; Gaps 7;
QY 9 KGLLALLPPGAA---STOYCTGDMKRLPASPETHLMDLRHLYOGCQVVOGNELELY 64
Db 8 WVVVSLVLAAGTVQPSDSQVACGETENKLSLSLEQQYRALRKYENGEVVMGNELEIS 67
QY 65 LPTNASTLSDIOEVOGVYLAHNOVROVPLQRLRIYVGTQLEFEDNVALAVLNDPRLN 124
Db 68 IEHNRLSLFRLKRYEYTVLALNOFRLPLENRLITIGTKLYEDRYALALFLNKRKG 127
QY 125 NTPPYTGAASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFRKNQALATLI 184
Db 128 NF-----GLQELGKLNLEILNGVYVNDKNKFLCADITIHMDIYRNPWPSLTLV 178
QY 185 DPNRSFACHPSCBCKSGKSCWGBSSSDCOSLTRVYACAGC-ARCKGRLPTDCCHEQCAAG 243
Db 179 STNGSSGCGCRKHSCTG-RCMGPTENHCQTLTRTYCAEQCDRCGRPYVSDCHRECAAG 237
QY 244 CNGPKHSDCLALHNHSGICELHCPALVTYNTDFEPMNDEGRYTFGASCVTACPVNY 303
Db 236 CGSPKDTDFACMNFENSDACVTCQCFVYNTTFQLEHNFNAKTYTAYGAFCKKCPHNF 297
QY 304 LSTDVSGCTVCPPLHNOEYTAEDGQRCBCKSPCAR-----GTHSL 346
Db 298 V-VDSSSCVRACPSKMEY-FENGIMCKPCPDICPKACDGIGTGLM 343

RESULT 10
A36223
kinase-related transforming protein (erbB3) (EC 2.7.1.1) precursor - human
C:Species: Homo sapiens (man)
C:Date: 04-Oct-1991 #sequence_revision 13-Jan-1993 #text_change 17-Nov-2000
C:Accession: A36223; I59164
R:Kraus, M.H.; Issing, W.; Miki, T.; Popescu, N.C.; Aaronson, S.A.

Proc. Natl. Acad. Sci. U.S.A. 86, 9193-9197, 1989
A:Title: Isolation and characterization of ERBB3, a third member of the ERBB/epiderma
A:Reference number: A36223; MUID:90083234
A:Accession: A36223
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1342 <KRA>
A:Cross-references: GB:M29366
R:Plowman, G.D.; Whitney, G.S.; Neubauer, M.G.; Green, J.M.; McDonald, V.L.; Todaro,
Proc. Natl. Acad. Sci. U.S.A. 87, 4905-4909, 1990
A:Title: Molecular cloning and expression of another epidermal growth factor receptor
A:Reference number: I59164; MUID:90311312
A:Accession: I59164
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-559, 'G', 561-957, 'F', 959-1063, 'G', 1065-1342 <RES>
A:Cross-references: GB:M34309; NID:9183990; PID:AAA3579.1; PID:9306841
C:Genetics:
A:Gene: GDB:ERBB3; HER3
A:Cross-references: GDB:119880; OMIM:190151
A:Map position: 12q13-12q13
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C:Keywords: ATP; phosphotransferase
F:707-972/Domain: protein kinase homology <KIN>
F:715-723/Region: protein kinase ATP-binding motif

Query Match 32.2%; Score 735.5; DB 2; Length 1342;
Best Local Similarity 44.0%; Pred. No. 9.5e-43;
Matches 150; Conservative 49; Mismatches 121; Indels 21; Gaps 9;
QY 10 GLLLALLPPGAA--STOYCTGDMKRLPASPETHLMDLRHLYOGCQVVOGNELELYPT 67
Db 11 GLFELANGSEVGNQANCPGLNGLSVTGAENQYQTLTKYECGEVVMGNELELVLTGH 70
QY 68 NASTSLFDIOEVOGVYLAHNOVROVPLQRLRIYVGTQLEFEDNVALAVLNDPRLNNTT 127
Db 71 NADLSFLQWIREYTVYLVAMNEFSTLPLRLRVAGTQVYDGKFAIFVM-----LANYT 125
QY 128 PYTGAASPGGLRELQRLSLTEILKGVLIQRPOLCYODTILMKDIFRKNQALATLI 187
Db 126 ---NSSHALROLRLTQLEILISGVYLEKNDKLCMHPTIDMRDIVRDRD---AEIVKD 178
QY 188 RSRACHPCSPMKGSRGCESESDCOSLTRVYACAGC-ARCKGRLPTDCCHEQCAAGCTG 246
Db 179 NRSRCPCHVCKG-RCMGPRSEDQTLTKITCAQCCHGCHGCRPNCCCHDECAAGSG 237
QY 247 PKHSDCLALHNHSGICELHCPALVTYNTDFEPMNDEGRYTFGASCVTACPVNYLST 306
Db 238 PDQTCFACRHNENDGACVPRCPQPLVYVKLTFQLEPDPHRTKYQYGVGCVASCPHNFV-V 296
QY 307 DVSGCTVCPPLHNOEYTAEDGQRCBCKSPCAR---GTHS 344
Db 297 DQTSVCVRACPPDKMEVD-KNGLKMEPCGGLCPKACGEGTS 336

RESULT 11
JC4387
epidermal growth factor receptor homolog precursor - rat
N:Alternate names: ErbB3 protein; HER3 protein
C:Species: Rattus norvegicus (Norway rat)
C:Date: 17-Jan-1996 #sequence_revision 19-Apr-1996 #text_change 13-Nov-1998
C:Accession: JC4387
R:Hellier, N.J.; Kim, H.H.; Greaves, C.H.; Sierke, S.L.; Koland, J.G.
Gene 165, 279-284, 1995
A:Title: Cloning of the rat ErbB3 cDNA and characterization of the recombinant protel
A:Reference number: JC4387; MUID:96096535
A:Accession: JC4387
A:Molecule type: mRNA
A:Residues: 1-1339 <HEL>
A:Cross-references: GB:U29339; NID:9915389; PID:9915390
A:Experimental source: liver
A:Note: The authors translated the codon AAC for residue 369 as Thr and GTP for resid

C:Comment: This protein is a functional heregulin receptor that transduces signals to the C:Genetics:

A:Gene: ErbB3

C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolog

C:Keywords: ATP; growth factor receptor; liver; phosphoprotein; transmembrane protein

F:1-19/Domain: signal sequence #status predicted <SIG>

F:20-1339/Product: epidermal growth factor homolog #status predicted <MAT>

F:640-659/Product: transmembrane #status predicted <TM>

F:705-970/Domain: protein kinase homology <KIN>

F:713-721/Region: protein kinase ATP-binding motif

F:939,1051,1156,1194,1219,1259,1273,1286,1325/Binding site: phosphate (Tyr) (C

Query Match 30.3%; Score 692; DB 2; Length 1339;

Best Local Similarity 41.8%; Pred. No. 9,1e-40;

Matches 146; Conservative 46; Mismatches 111; Indels 26; Gaps 10;

Db 3 LAALCRWGLLALLPPGAA---STQVCTGTDMLRLPASPETHLMDLRHLYOGCCOVVGN 59

Db 7 LQVLC---FLLSLARGSEMGNSQAVCPGTLNGLSVTGDAQNQYQTLVKYKECEVWGN 62

Db 60 LEITYLPNASTSLFDIOIEVGYVLLAHNOVROYPLQRLRIVRGTOLEFEDNYALAVLDN 119

Db 63 LEIVLTGHMADLSFLQWIREVATAYLVAMNEFSYLPENLIVVRGTQYDCKFAIFVW-- 120

Db 120 GDPINNTPTVPGASPGGLREQLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNQL 179

Db 121 ---LNYNT---NSSHALRQLKFTQLEILSGVYIEKNDLCHMDTIDMDIYRVR-- 170

Db 180 ALTLIDNRSRACHPCSPCKSGKRCWGESSEDCOSLRTVCAGGC-ARCKGRLPTDCHE 238

Db 171 GAELIVKKNGANCPCHPCHEVCKG-RCMGPGPDCCQLTKTICAPQCNCGKFCFBNPQCCHD 229

Db 239 QCAAGCTGPKHSDDLACLHFNHSGICELHCPALVYNTDTFESMPNPGRTFGASCVTA 298

Db 230 ECAGGCGSPQDTDFACFRFNDSCACVPCPEPLVYNNKLTFLQLEPNPTKYQGVCAVS 289

Db 299 CPYVNLSTDVGSCTLVCPHLHNOEYTAEDGTQRCCKSPCAR---GTHS 344

Db 290 CPNHFV-VDQTFVCYRACPPDKMEVD-KHGLMKCEPGGLCPKACBGTGS 336

RESULT 12

S06142

protein-tyrosine kinase (EC 2.7.1.112) mrk-y precursor - southern platyfish

N:Alternate names: epidermal growth factor receptor homolog; kinase-related transforming

C:Species: Xiphophorus maculatus (southern platyfish)

C>Date: 10-Sep-1999 #sequence-revision 10-Sep-1999 #text-change 18-Feb-2000

C:Accession: S06142; S13809

R:Wiltbrodt, J.; Adam, D.; Maltzschek, B.; Maeuener, W.; Raulf, F.; Telling, A.; Roberts

Nature 341, 415-421, 1989

A:Title: Novel putative receptor tyrosine kinase encoded by the melanoma-inducing Tu loc

A:Reference number: S06142; MUID:90015140

A:Accession: S06142

A:Molecule type: DNA

A:Residues: 1-1166 <MIT>

A:Cross-references: EMBL:X16891; NID:965290; PIDN:CAA3770.1; PID:965291

R:Adam, D.; Maeuener, W.; Scharf, M.

Oncogene 6, 73-80, 1991

A:Title: Transcriptional activation of the melanoma inducing Xmrk oncogene in Xiphophorus

A:Reference number: S13807; MUID:9112882

A:Accession: S13809

A>Status: preliminary; translation not shown

A:Molecule type: DNA

A:Residues: 821-1025, 'N', 1027-1098, 'A', 1100-1166 <ADA>

A:Cross-references: EMBL:X56319; NID:965284; PIDN:CAA39763.1; PID:965285

C:Genetics:

A:Gene: mrk

A:Map position: Y

A:Introns: 872/3; 898/1; 947/1; 979/3; 1025/3; 1056/1

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor; phosphotransferase; transmembrane protein; tyrc

F:1-25/Domain: signal sequence #status predicted <SIG>

F:26-1166/Product: kinase-related transforming protein (Tu) #status predicted <MAT>

F:707-972/Domain: protein kinase homology <KIN>

F:715-723/Region: protein kinase ATP-binding motif

Query Match 29.8%; Score 681.5; DB 1; Length 1166;

Best Local Similarity 42.4%; Pred. No. 4,1e-39;

Matches 145; Conservative 46; Mismatches 136; Indels 15; Gaps 8;

Db 4 AALCRWGLLALLPPGAAT---QVCTGTDMLRLPASPETHLMDLRHLYOGCCOVVGN 59

Db 8 AALLQ---LLVLISIRCCSTDPDRKVCQGTSMQTM--LDNHYLKKMKMTSGCNVLEN 62

Db 60 LEITYLPNASTSLFDIOIEVGYVLLAHNOVROYPLQRLRIVRGTOLEFEDNYALAVLDN 119

Db 63 LEITYTOENQSLFLOSIQEVGYVLLAMNEFSYLPENLIVVRGTQYDCKFAIFVW-- 122

Db 120 GDPINNTPTVPGASPGGLREQLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNQL 179

Db 123 YQK-NPSP--DYQVGLKQLSLNLEILSGVKVSHNPLCNVETIMMDIVDKTSNP 179

Db 180 ALTLIDNRSRACHPCSPCKSGKRCWGESSEDCOSLRTVCAGGC-ARCKGRLPTDCHE 238

Db 180 TMLNIPHAEEQCKDHCVCNAGSCWAPGHCOKFTKLCAEQCNRCRGPPIKIDCNE 239

Db 239 QCAAGCTGPKHSDDLACLHFNHSGICELHCPALVYNTDTFESMPNPGRTFGASCVTA 298

Db 240 HCAGGCTGPRATDCLACDFNDGCKDTPPPKTYIDVSHQVVDNPNIKYTFGACVKE 299

Db 299 CPYVNLSTDVGSCTLVCPHLHNOEYTAEDGTQRCCKSPCAR 340

Db 300 CPNHYVTE-GACVRCSCAGMLEVD-ENGRKSCRPDCGVCK 339

RESULT 13

A27131

epidermal growth factor receptor - fruit fly (Drosophila melanogaster) (fragment)

C:Species: Drosophila melanogaster

C>Date: 19-Nov-1988 #sequence-revision 19-Nov-1988 #text-change 23-May-1997

C:Accession: A27131

R:Schejter, E.D.; Segal, D.; Glazer, L.; Shilo, B. Z.

Cell 46, 1091-1101, 1986

A:Title: Alternative 5' exons and tissue-specific expression of the Drosophila EGF re

A:Reference number: A27131; MUID:87002474

A:Accession: A27131

A:Molecule type: mRNA

A:Residues: 1-843 <SCH>

C:Genetics:

A:Gene: FlyBase:Egfr

A:Cross-references: FlyBase:FBgn0003731

C:Superfamily: epidermal growth factor receptor; protein kinase homology

C:Keywords: ATP; growth factor receptor

Query Match 25.2%; Score 575.5; DB 2; Length 843;

Best Local Similarity 36.6%; Pred. No. 5,4e-32;

Matches 119; Conservative 45; Mismatches 130; Indels 31; Gaps 7;

Db 24 QVCTGTDMLRLPASPETHLMDLRHLYOGCCOVVGNLETLPLT-NASTSLFDIOIEVYG 82

Db 51 KVCIGTNSRLSPVSKNHNKLNLDRTNCTYVGNGLKLTMLPNEINDLSFLDNIREVTG 110

Db 83 YVLIANNOVROYPLQRLRIVRGTOLEFEDNYALAVLDNGBPLNNTPTVPGASPGGL 137

Db 111 YILISHVDKVKVPPKLOIIRGRFLFSLSVEEKYALFV-----TVSKM 154

Db 138 RELDRLSLTEILKGVLIQRNPOLCYODTILMKDIFHKNNQALTLIDNRSRACHPCSP 197

Db 155 YTELPDRLDVLNGQVGFHNHNYNLCHMRTIQWSEIVSNGTDAYVNYDTAPRCPCKHE 214

Db 198 MCKSGRCWGESSEDCOSLRTVCAGGCA--RCKGRLPTDCHEQCAAGCTGPKHSDDLAC 255

Db 215 SCTHG-CMGSEPKKCKQKFSKLTGSPQCAAGRCVYRPKPRECHNLRCAGGCTGPTQKDCIAC 273

OY 256 LHFNSGICELCPALVYNTDTFESMPNREGRTYEGASCATACPNYLSTDVSCCLYC 315
DB 274 KNFPEANVSKKECCPPMKRYNPTTYLETNPSGKAIYGAICVKECP-CHILRDNGACVNSC 332
OY 316 PLHNOEVTAEADGTORCEKCSKPCAR 340
DB 333 PQDKMDKGE-----CVPNGPCPK 352
RESULT 14
S70712
protein-tyrosine kinase (EC 2.7.1.112) let-23 precursor - Caenorhabditis elegans
N/Alternate names: receptor tyrosine kinase let-23
C/Species: Caenorhabditis elegans
C/Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 17-Nov-2000
C/Accession: S70712; S73101; S13422; T27682
R/Sakai, T.; Koga, M.; Ohshima, Y.
J. Mol. Biol. 256, 548-555, 1996
A/Title: Genomic structure and 5' regulatory regions of the let-23 gene in the nematode
A/Reference number: S70712; MUID:96177760
A/Accession: S70712
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1374 <SAK>
A/Cross-references: EMBL:D63426
A/Experimental source: strain N2
R/Koga, M.
submitted to the EMBL Data Library, July 1995
A/Reference number: S73101
A/Accession: S73101
A/Molecule type: DNA
A/Residues: 1-50, 'G', 52-1374 <KOG>
A/Cross-references: EMBL:D63426; NID:g1407562; PIDN:BA009729.1; PID:d1010375; PID:g14075
A/Experimental source: strain N2
R/Ariolan, R.V.; Koga, M.; Mendel, J.E.; Ohshima, Y.; Sternberg, P.W.
Nature 348, 693-699, 1990
A/Title: The let-23 gene necessary for Caenorhabditis elegans vulval induction encodes a
A/Reference number: S13422; MUID:91080919
A/Accession: S13422
A/Molecule type: mRNA
A/Residues: 52-1374 <ARO>
R/Thomas, K.
submitted to the EMBL Data Library, March 1996
A/Reference number: 220404
A/Accession: T27682
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 52-1374 <MIL>
A/Cross-references: EMBL:270038; PIDN:CAA93882.1; GSPDB:GN00020; CESP:ZK1067.1
A/Experimental source: clone ZK1067
C/Genetics:
A/Gene: let-23; CESP:ZK1067.1
A/Map position: 2
A/Introns: 44/1; 51/1; 90/1; 112/3; 165/3; 217/1; 290/1; 379/1; 418/1; 448/2; 565/1; 608
C/Suprafamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase homolo
C/Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein ki
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1374/Product: protein-tyrosine kinase let-23 #status predicted <MAT>
F:934-1199/Domain: protein kinase homology <KIN>
F:942-950/Region: protein kinase ATP-binding motif

Query Match 18.3%; Score 419; DB 2; Length 1374;
Best Local Similarity 29.0%; Pred. No. 4.8e-21;
Matches 102; Conservative 59; Mismatches 131; Indels 60; Gaps 14;

OY 25 VCTGDMKRLRPASPEHLDMLRHLRYOCQVQGNLELTYPNTN----- 68
DB 90 LCGTNGISRGCTGNI-LEDEFTMYRGCRRYGNGLEITWIANELKKRSTNSTVDPK 148
OY 69 -----ASLSFLQDIOEVQGVYLIANNOVROYPLQRLRVGTQLEFDNYALAVLDNGDP 122

DB 149 NEDSPKSIENFDNLEIRGSLIIYRANIOKISPRRLRVIGDEVEFDN-ALYIHKNDK- 206
OY 123 LNNTPPVYASPGGLRELOLSIFTEILKGVLIQRNPOLC-ODTIIMKLOFHNNOLAL 181
DB 207 -----VHEVVMRELRYIRNGSVTIDONPMCYIGKIDMKELLYDPD-VQ 250
OY 182 TLIDTNRSRACH-----PCSPMKSGRCGSESESDOSLTRVACGAGCARC--KGPL 231
DB 251 KYETTNSHOCYQNKSMNAKCHESC-NDKCWSGSDNDQORYRYSVCPSKSCQCFYNSNTS 309
OY 232 PTDCCHEQCAAGCTGPKHSDCLAHFNHSGICELHCPALVYNTDTFESMPNREGRTYE 291
DB 310 SYECDSDACLGCTGHGPKNCIACSKYELDICTECPSRKIFHNKTRGLVFENPDGRYON 369
OY 292 GASCVTACPNYVL-STDVGSCITVC-PLHNOEVTAEADGTORCEK-SKPCAR 340
DB 370 GNCVKCECPPELLIENDV--CVRHCSGDGHYDATKD--VRECEKCRSSCPK 417
RESULT 15
S70713
protein-tyrosine kinase let-23 precursor homolog - Caenorhabditis vulgaris
N/Alternate names: receptor tyrosine kinase let-23 homolog
C/Species: Caenorhabditis vulgaris
C/Date: 21-Apr-1997 #sequence_revision 09-May-1997 #text_change 19-Dec-1997
R/Sakai, T.; Koga, M.; Ohshima, Y.
J. Mol. Biol. 256, 548-555, 1996
A/Title: Genomic structure and 5' regulatory regions of the let-23 gene in the nemato
A/Reference number: S70712; MUID:96177760
A/Accession: S70713
A/Status: nucleic acid sequence not shown
A/Molecule type: DNA
A/Residues: 1-1369 <SAK>
A/Cross-references: EMBL:D63427
C/Genetics:
A/Gene: let-23
A/Introns: 42/1; 49/1; 83/1; 105/3; 155/3; 207/1; 280/1; 369/1; 408/1; 438/2; 555/1;
C/Suprafamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase hom
C/Keywords: ATP; phosphotransferase; transmembrane protein; tyrosine-specific protein
F:1-28/Domain: signal sequence #status predicted <SIG>
F:29-1369/Product: protein-tyrosine kinase let-23 homolog #status predicted <MAT>
F:929-1194/Domain: protein kinase homology <KIN>
F:937-945/Region: protein kinase ATP-binding motif

Query Match 18.2%; Score 416; DB 2; Length 1369;
Best Local Similarity 28.6%; Pred. No. 7.7e-21;
Matches 105; Conservative 59; Mismatches 135; Indels 68; Gaps 13;

OY 25 VCTGDMKRLRPASPEHLDMLRHLRYOCQVQGNLELTYPNTN----- 68
DB 83 VCSGNNLSLRKSGSNT-LEDEHMYRRCRVRVYNGLEITWIANELIOKWRSTNOTVDAD 141
OY 69 -----ASLSFLQDIOEVQGVYLIANNOVROYPLQRLRVGTQLEFDNYALAVLDNGDPLN 124
DB 142 IDYLKTVNFPDLHERISLIIYRANIOKISPRKLRVIGDEVFHDN-SLYIHQHEK--- 197
OY 125 NTPPTGASPGRLQRLSLTEILKGVLIQRNPOLCYOT-ILMKDIFPKNNQALTL 183
DB 198 -----VNLMVKLELRVINGSVSIQNNRPMFLTKDWMNILLXDSRQKVE- 244
OY 184 IDTNRSRACHPCSPV-----CKGSRGWSSESDOSLTRVACGAGCARCKGPLPT--- 233
DB 245 -XXNSHKACWANGELIASXHEKCK-DKCGWGSDNDQORYRYSVCPSKSCQCFYNSITDSY 302
OY 234 DCHHEQCAAGCTGPKHSDCLAHFNHSGICELHCPALVYNTDTFESMPNREGRTYFGA 293
DB 303 EECDSGSCGCGCNGHNRDSCIASCKYEMDMCIDIOPKARKINHHKTRGLRVPRDGRYQNGN 362
OY 294 SCVTACPNYVLSTDVGSCITVCPLHNOEVTAEADGT---ORCEK-SKPCARGTSHSLPRP 349
DB 363 HCVCKECPPELLIXND-----VCVRHCSGEGHHYDATKDMRECEKCPSSSC-----PKI 409

OY 350 AAVVPL 356
| |
Db 410 CTVDGPL 416

Search completed: April 11, 2002, 09:30:11
Job time: 108 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:34:18 ; Search time 15.52 Seconds

(Without alignments)
989.856 Million cell updates/sec

Title: US-09-234-208b-2

Perfect score: 2287

Sequence: 1 MELALCRMGILLALLPFGA.....VGKGPDAHVAVNLRYEG 419

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 3664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	1255	1	ERB2_HUMAN
2	1608.5	70.3	1257	1	ERB2_RAT
3	1571	68.7	1254	1	ERB2_MESAU
4	796.5	34.8	703	1	EGFR_CHICK
5	793	34.7	1210	1	EGFR_HUMAN
6	789	34.5	1210	1	EGFR_MOUSE
7	775	33.9	1308	1	ERB4_HUMAN
8	735.5	32.2	1342	1	ERB3_HUMAN
9	698	30.5	1339	1	ERB3_RAT
10	681.5	29.8	1167	1	XMRK_XIPMA
11	574.5	25.1	1426	1	EGFR_DROME
12	419	18.3	1323	1	LT23_CAEEL
13	342.5	15.0	1363	1	ILPR_BRLA
14	294	12.9	2146	1	INSR_DROME
15	291	12.7	1477	1	HTK7_HYDAT
16	278	12.2	1300	1	IRR_CAVPO
17	270	11.8	581	1	IRR_RAT
18	269.5	11.8	1607	1	MIPR_LYMST
19	269	11.8	1297	1	IRR_HUMAN
20	264.5	11.6	1382	1	INSR_HUMAN
21	263.5	11.5	1383	1	INSR_RAT
22	261	11.4	1372	1	INSR_MOUSE
23	258	11.3	1390	1	INSR_AEDAE
24	257.5	11.3	1367	1	IGIR_HUMAN
25	252.5	11.0	1370	1	IGIR_RAT
26	250.5	11.0	1373	1	IGIR_MOUSE
27	169	7.4	1696	1	PKCS_BRACL
28	144	6.3	1877	1	PKCS_MOUSE
29	142	6.2	1959	1	AGRI_RAT
30	131.5	5.7	913	1	PK5_HUMAN
31	129.5	5.7	1680	1	FUR2_DROME
32	124.5	5.4	1877	1	PK5_RAT
33	123.5	5.4	937	1	PAC4_RAT

34	123	5.4	417	1	WSL1_HUMAN	Q93038 h wsl-1 pro
35	119.5	5.2	667	1	TS11_GIALA	003185 giardia lam
36	119	5.2	2248	1	ZAN_RABIT	P57999 oryctolagus
37	118.5	5.2	551	1	LEM2_RABIT	P27113 oryctolagus
38	118.5	5.2	3635	1	LMAS_MOUSE	061001 mus musculu
39	117	5.1	469	1	PROP_HUMAN	P27918 homo sapien
40	117	5.1	3075	1	LMAL_HUMAN	P25391 homo sapien
41	116.5	5.1	969	1	PAC4_HUMAN	P29132 homo sapien
42	116.5	5.1	1173	1	TSP1_XENLA	P35448 xenopus lae
43	115	5.0	3106	1	LMAS_MOUSE	060675 mus musculu
44	114.5	5.0	484	1	LEM2_PIG	P98110 sus scrofa
45	113	4.9	1122	1	TIE2_MOUSE	Q02858 mus musculu

ALIGNMENTS

RESULT	ID	ERB2_HUMAN	STANDARD:	PRT:	1255 AA.
AC	P04626	13-AUG-1987 (Rel. 05, Created)			
DT	13-AUG-1987 (Rel. 05, Last sequence update)				
DT	20-AUG-2001 (Rel. 40, Last annotation update)				
DE	RECEPTOR PROTEIN-TYROSINE KINASE ERBB-2 PRECURSOR (EC 2.7.1.112)				
DE	(P185ERB2) (NEU PROTO-ONCOGENE) (C-ERBB-2) (TYROSINE KINASE-TYPE CELL SURFACE RECEPTOR HER2) (MLN 19).				
DE	ERBB2 OR HER2 OR NGL OR NEU.				
GN	Homo sapiens (Human).				
OS	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.				
OX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=86118663; PubMed=3003577;				
RA	Yamamoto T., Ikawa S., Akiyama T., Semba K., Nomura N., Miyajima N.,				
RA	Saito T., Toyoshima K.;				
RT	"Similarity of protein encoded by the human c-erb-B-2 gene to				
RT	epidermal growth factor receptor.";				
RL	Nature 319:230-234(1986).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=86070181; PubMed=2999974;				
RA	Coussens L., Yang-Feng T.L., Liao Y.C., Chen E., Gray A.,				
RA	McCrath J., Seeburg P.H., Libermann T.A., Schlessinger J.,				
RA	Francie U., Levinson A., Ullrich A.;				
RT	"Tyrosine kinase receptor with extensive homology to EGF receptor				
RT	shares chromosomal location with neu oncogene.";				
RL	Science 230:1132-1139(1985).				
RN	[3]				
RP	SEQUENCE OF 737-1031 FROM N.A.				
RX	MEDLINE=86016729; PubMed=2959367;				
RA	Semba K., Kanata N., Toyoshima K., Yamamoto T.;				
RA	"A v-erbB-related protooncogene, c-erbB-2, is distinct from the				
RT	c-erbB-1/epidermal growth factor receptor gene and is amplified in a				
RT	human salivary gland adenocarcinoma.";				
RL	Proc. Natl. Acad. Sci. U.S.A. 82:6497-6501(1985).				
RN	[4]				
RP	VARIANTS VAL-654 AND VAL-655.				
RX	MEDLINE=93194196; PubMed=8095488;				
RA	Eheant A., Low J., Wallace R.B., Wu A.M.;				
RT	"Characterization of a new allele of the human ERBB2 gene by allele-				
RT	specific competition hybridization.";				
RL	Genomics 15:426-429(1993).				
CC	-1- FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,				
CC	ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A				
CC	POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-				
CC	ALPHA AND AMPHIREGULIN.				
CC	-1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +				
CC	PROTEIN TYROSINE PHOSPHATE.				
CC	-1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS				
CC	(POTENTIAL).				
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.				

CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES (BY SIMILARITY).
CC -1- POLYMORPHISM: THERE ARE FOUR ALLELES DUE TO THE VARIATIONS IN
CC POSITIONS 654 AND 655. ALLELE B1 (654-ILE-ILE-655) HAS A FREQUENCY
CC OF 0.782; ALLELE B2 (654-ILE-VAL-655) HAS A FREQUENCY OF 0.206;
CC ALLELE B3 (654-VAL-VAL-655) HAS A FREQUENCY OF 0.012.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M11767; AAA35808.1; JOINED.
DR EMBL; M11761; AAA35808.1; JOINED.
DR EMBL; M11762; AAA35808.1; JOINED.
DR EMBL; M11763; AAA35808.1; JOINED.
DR EMBL; M11764; AAA35808.1; JOINED.
DR EMBL; M11765; AAA35808.1; JOINED.
DR EMBL; M11766; AAA35808.1; JOINED.
DR EMBL; M11730; AAA35493.1; -
DR EMBL; M12036; AAA35978.1; -
DR EMBL; X03363; CAA27060.1; -
DR PIR; A25491; A25491.
DR PIR; A24571; A24571.
DR HSSP; P11362; IFE1.
DR MIM; 164870; -
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam; PF00757; Furin-like; 1.
DR Pfam; PF00069; Pkinase; 1.
DR Pfam; PF01030; Recep_L_domain; 2.
DR SMART; SMO0261; FU; 3.
DR SMART; SMO0219; TyrcKc; 1.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Polymorphism.
FT SIGNAL 1 21
FT CHAIN 22 1255
FT DOMAIN 22 652
FT TRANSMEM 653 675
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT MOD_RES 1139 1139
FT MOD_RES 1248 1248
FT CARBOHYD 68 68
FT CARBOHYD 124 124
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 654 654
FT VARIANT 655 655
FT VARIANT 655 655
FT CONFLICT 1170 1170
FT SEQUENCE 1255 AA; 137909 MW; 39B9DFDA04DC962 CRC64;
Query Match 82.1%; Score 1878; DB 1; Length 1255;
Best Local Similarity 83.0%; Pred. No. 3e-139;

Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
QY 1 MEIALACRWGLLLALPRGASTOVCTGDMKLRLPASPEHLMRLHYGCGVGNL 60
DB 1 MELALACRWGLLLALPRGASTOVCTGDMKLRLPASPEHLMRLHYGCGVGNL 60
QY 61 ELTYLPTNASLFLQIDQVGVYLIANOVROYPLRLRIVRGTOLEFEDNYVALVDNG 120
DB 61 ELTYLPTNASLFLQIDQVGVYLIANOVROYPLRLRIVRGTOLEFEDNYVALVDNG 120
QY 121 DPLNTPVPGASPGEGLEQLRSLTEILKGVLIQRNPOLCYODTILMKDIFRKNOLA 180
DB 121 DPLNTPVPGASPGEGLEQLRSLTEILKGVLIQRNPOLCYODTILMKDIFRKNOLA 180
QY 181 LTLIDNRSRCHPCSPCKSCRCWGESSECCOSTLRVCAGGARCGPLPTCCHEOC 240
DB 181 LTLIDNRSRCHPCSPCKSCRCWGESSECCOSTLRVCAGGARCGPLPTCCHEOC 240
QY 241 AAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDFEESMPNDEGRYTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLAFHNSGICELHCPALVTYNTDFEESMPNDEGRYTFGASCVTACP 300
QY 301 YNYLSTDVSGCTLVCPHLNDEVTAEDGTQRCCKSPCAR-----GTHSLPRPAAVVP 355
DB 301 YNYLSTDVSGCTLVCPHLNDEVTAEDGTQRCCKSPCARCYGLGMEHLREVAVTSAN 360
QY 356 LRMFG--PAHPVLSFLRPSMDLVSAFSLPLAPLSPSVPI-----SPVSGRGP 405
DB 361 IQERAGCKKIRGSLAFLESDGDPASNT---APLQEPQLOVFETLEITGYLYISAMPD 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 SLPLDSVQNQLQVIRG 433
RESULT 2
ID ERB2_RAT STANDARD; PRT; 1257 AA.
AC P06494;
DT 01-JAN-1988 (Rel. 06, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ERB2-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
DE (P185ERB2) (NEU PROTO-ONCOGENE) (EPIDERMAL GROWTH FACTOR RECEPTOR-RELATED PROTEIN).
GN ERB2 OR NEU
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroblastoma; PubMed=3945311;
RX BARGMANN C.I., HUNG M.-C., WEINBERG R.A.;
RT "The neu oncogene encodes an epidermal growth factor receptor-related protein";
RL Nature 319:226-230(1986).
RN [2]
RP SEQUENCE OF 852-905 FROM N.A.
RC TISSUE=Sciatic nerve;
RX MEDLINE=91222560; PubMed=2025425;
RA Lai C., Lemke G.;
RT "An extended family of protein-tyrosine kinase genes differentially expressed in the vertebrate nervous system";
RL Neuron 6:691-704(1991).
RN [3]
RP STRUCTURE BY NMR OF 650-668.
RX MEDLINE=92155181; PubMed=1346763;
RA Gullick W.J., Bottomley A.C., Lofts F.J., Doak D.G., Mulvey D.,
RA Newman R., Crumpton M.J., Sternberg M.J.E., Campbell I.D.;
RT "Three dimensional structure of the transmembrane region of the proto-oncogenic and oncogenic forms of the neu protein.";

RL EMBL J. 11:43-48(1992).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN.
 CC - CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS.
 CC THE CONSTITUTIVELY ACTIVATED ONCOGENIC VARIANT FORMS A HOMODIMER.
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES (BY SIMILARITY).
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, X03362; CAA27059.1; ALT_INT.
 DR PIR, A24562; TVRTNU.
 DR HSSE, P11362; 1FGI.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TyrKc; 1.
 DR PROSITE, PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE, PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE, PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
 KW Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
 KW Proto-oncogene; Disease mutation.
 FT SIGNAL 1 21
 FT CHAIN 22 1257
 FT DOMAIN 22 654
 FT TRANSMEM 655 677
 FT DOMAIN 678 1257
 FT DOMAIN 159 369
 FT DOMAIN 473 646
 FT DOMAIN 722 989
 FT NP_BIND 728 736
 FT BINDING 755 755
 FT ACT_SITE 847 847
 FT MOD_RES 1141 1141
 FT MOD_RES 1250 1250
 FT CARBOHYD 68 68
 FT CARBOHYD 188 188
 FT CARBOHYD 260 260
 FT CARBOHYD 532 532
 FT CARBOHYD 573 573
 FT CARBOHYD 631 631
 FT VARIANT 661 661
 FT SEQUENCE 1257 AA; 138831 MW; 6129264583011402 CRC64;
 Query Match 70.3%; Score 1608.5; DB 1; Length 1257;
 Best Local Similarity 85.0%; Pred. No. 36e-118;
 Matches 294; Conservative 15; Mismatches 36; Indels 1; Gaps 1;
 Oy 1 MELALACRMGLIALLPFGASTOVCTGDMKRLPASPETHLMDLRHLVYGCQVVGNTL 60
 Db 1 MELAMCMRGFLALPLPGINGTGYCTGDMKRLPASETHLMDLRHLVYGCQVVGNTL 60
 Oy 61 ELTYLPANASLSFLDIOEVGYVLIANQVQVPLQRLIRIVRGTOLEEDNALAVLDNG 120
 Db 61 ELTYLPANASLSFLDIOEVGYVLIANQVQVPLQRLIRIVRGTOLEEDNALAVLDNG 120

Db 61 ELTYLPANASLSFLDIOEVGYVLIANQVQVPLQRLIRIVRGTOLEEDKAYALAVLDNR 120
 Oy 121 DLINMTTPVT-GASPGRLBELQSLSTELTKGVLQIQRNPOLCYODTIMKDFKNNOL 179
 Db 121 DPQDNVAASTPRTPEGLNELQSLSTELTKGVLIRGPPOLCYODIMVLMKVFERNOL 180
 Oy 180 ALTLIDNRSRACHPCSPKCKSGRCGSESEDCQSLTFTVCAGCARGKPLPTDCHEQ 239
 Db 181 AVVDIDNRSRACHPCAPACKKNHMGESPECCQLTGTICSGCARGKRLPTDCHEQ 240
 Oy 240 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDFTESMPNDEGRYTFGASCVTAC 299
 Db 241 CAAGCTGPKHSDCLACLHFNHSGICELHCPALVYNTDFTESMNPDEGRYTFGASCVTAC 300
 Oy 300 PNYNTSTVDSCTVCPILHNOEVTADGROREKSKPCARQTHSL 345
 Db 301 PNYNTSTVDSCTVCPILHNOEVTADGROREKSKPCARQVGYL 346
 RESULT 3
 ID ERB2_MESAU STANDARD; PRT; 1254 AA.
 AC 060553;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE ERBB-2 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
 DE (P185ERBB2) (NEU PROTO-ONCOGENE).
 GN ERBB2 OR NEU.
 OS Mesocricetus auratus (Golden hamster).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;
 OC Mesocricetus.
 NCBI_Taxid=10036;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Nerve;
 RX MEDLINE=94193007; Pubmed=7908275;
 RA Nakamura T., Ushijima T., Ishizaka Y., Nagao M., Arai M.,
 RA Yamazaki Y., Ishikawa T.;
 RL Gene 140:251-255(1994).
 CC - FUNCTION: ESSENTIAL COMPONENT OF A NEUREGULIN-RECEPTOR COMPLEX,
 CC ALTHOUGH NEUREGULINS DO NOT INTERACT WITH IT ALONE. GP30 IS A
 CC POTENTIAL LIGAND FOR THIS RECEPTOR. NOT ACTIVATED BY EGF, TGF-
 CC ALPHA AND AMPHIREGULIN (BY SIMILARITY).
 CC - CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC - SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
 CC (POTENTIAL).
 CC - SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC - PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
 CC RESIDUES.
 CC - SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL, D16295; BAA03801.1; -
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TyrKc; 1.

DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KM Transmembrane; Glycoprotein; Multigene family; Receptor; Signal.
KM Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Proto-oncogene; Disease mutation.
FT SIGNAL 1 21
FT CHAIN 22 1254
FT TRANSMEM 22 652
FT TRANSMEM 653 675
FT DOMAIN 676 1254
FT DOMAIN 158 368
FT DOMAIN 472 644
FT DOMAIN 720 987
FT NP_BIND 726 734
FT BINDING 753 753
FT ACT_SITE 845 845
FT MOD_RES 1139 1139
FT MOD_RES 1247 1247
FT CARBOHYD 68 68
FT CARBOHYD 125 125
FT CARBOHYD 187 187
FT CARBOHYD 259 259
FT CARBOHYD 530 530
FT CARBOHYD 571 571
FT CARBOHYD 629 629
FT VARIANT 658 658
FT VARIANT 659 659
SQ SEQUENCE 1254 AA; 138252 MW; 974C3791C21F2B81 CRC64;

Query Match 68.7%; Score 1571; DB 1; Length 1254;
Best Local Similarity 74.4%; Pred. No. 3.1e-115;
Matches 299; Conservative 26; Mismatches 67; Indels 10; Gaps 3;

QY 1 MELALCRWGLLALLPGAASTQVGTDMKRLPASEPETHLDMRLHYGCGVVGML 60
DB 1 MELAAWCGWGLLALLPGAASTQVGTDMKRLPASEPETHLDMRLHYGCGVVGML 60
QY 61 ELTYLPNTNLSFLDIOIOVGVLLAHNOVQVPLQRLRYRGQLFEDNVALAVLDNG 120
DB 61 ELTYLPNTNLSFLDIOIOVGVLLAHNOVQVPLQRLRYRGQLFEDNVALAVLDNR 120
QY 121 DPLNNTPTVGTGASPGGLRLQLRLTEILKGVLIORNLQYODITLKKDIFKKNQQA 180
DB 121 DPLNNTPTVGTGASPGGLRLQLRLTEILKGVLIORNLQYODITLKKDIFKKNQQA 180
QY 121 DPLNNTPTVGTGASPGGLRLQLRLTEILKGVLIORNLQYODITLKKDIFKKNQQA 180
DB 121 DPLNNTPTVGTGASPGGLRLQLRLTEILKGVLIORNLQYODITLKKDIFKKNQQA 180
QY 181 LTLIDTNRSPACHPCSPCKSGSRWGESSEDCOSLTRIVCAGGACARCKGPLTDCHEQC 240
DB 181 LTLIDTNRSPACHPCSPCKSGSRWGESSEDCOSLTRIVCAGGACARCKGPLTDCHEQC 240
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTGASCTTAC 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTGASCTTAC 300
QY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTGASCTTAC 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVYNTDFESMPNPGRTGASCTTAC 300
QY 301 YVLTSDVGSCTLVCPVLHNOVEYTAEDGTORCEKSKPCAR-----GTHSLLRPAVAVP 355
DB 301 YVLTSDVGSCTLVCPVLHNOVEYTAEDGTORCEKSKPCAR-----GTHSLLRPAVAVP 355
QY 301 YVLTSDVGSCTLVCPVLHNOVEYTAEDGTORCEKSKPCAR-----GTHSLLRPAVAVP 355
DB 301 YVLTSDVGSCTLVCPVLHNOVEYTAEDGTORCEKSKPCAR-----GTHSLLRPAVAVP 355
QY 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLLPLAPISPTSVPT 395
DB 356 LRMQPG--PAHPVLSFLRPSMDLVSAFYSLLPLAPISPTSVPT 395
QY 361 IOFAGCKKIFGSLAFLPESFD---GNPSSGIAPLTPPEQLQV 399
DB 361 IOFAGCKKIFGSLAFLPESFD---GNPSSGIAPLTPPEQLQV 399

RESULT 4
EGFR_CHICK STANDARD; PRT; 703 AA.
AC EGFR_CHICK P13387;
DT 01-JAN-1990 (Rel. 13. Created)
DT 01-JAN-1990 (Rel. 13. Last sequence update)
DT 20-AUG-2001 (Rel. 40. Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112) (CER (FRAGMENT)).

GN EGFR.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauromorpha; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88261272; PubMed=3260329;
RA Lax I., Johnson A., Hawk R., Sap J., Bellot F., Winkler M.,
RT Ullrich A., Vennstrom B., Schlessinger J., Givol D.;
RT "Chicken epidermal growth factor (EGF) receptor: cDNA cloning,
RT expression in mouse cells, and differential binding of EGF and
RT transforming growth factor alpha."
RL Mol. Cell. Biol. 8:1970-1978(1988).
CC -!- FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
CC AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
CC VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -!- MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND
CC CELL PROLIFERATION.
CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: M20386; AAA48760.1; -.
DR Interpro: IPR000494; EGFR_L.
DR Interpro: IPR000719; Euk_pkinase.
DR Interpro: IPR002174; Furin-like.
DR Interpro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 4.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; PARTIAL.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; PARTIAL.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; PARTIAL.
KM Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
FT SIGNAL 1 30
FT CHAIN 31 703
FT DOMAIN 31 654
FT TRANSMEM 655 667
FT TRANSMEM 668 703
FT DOMAIN 134 134
FT CARBOHYD 190 190
FT CARBOHYD 200 200
FT CARBOHYD 359 359
FT CARBOHYD 368 368
FT CARBOHYD 420 420
FT CARBOHYD 573 573
FT CARBOHYD 578 578
FT CARBOHYD 613 613
FT CARBOHYD 633 633
FT CARBOHYD 648 648
FT NON_TER 703
SQ SEQUENCE 703 AA; 77427 MW; AEF2DE1B735A690 CRC64;

Query Match 34.8%; Score 796.5; DB 1; Length 703;
Best Local Similarity 45.5%; Pred. No. 6.3e-55;
Matches 155; Conservative 55; Mismatches 112; Indels 19; Gaps 7;

QY 8 RWGLLALLPGAA-----STQVCTGDMKRLPASEPETHLDMRLHYGCGVVGML 61

```

Db      13 RGAAYLVLLLLGVALCSAVEEKKVCGGTNNKLTQLGHHVEDHFTSLQRMNCEVYLSNLE 72
Qy      62 LTYLPTNMSLSLTDIOEQGVYLAHNOYRVPRLRIYVGTOLFEENYALAVLNDG 121
Db      73 IYVYHNDLFTLKTIOEYAGVYLAHNVDPLENIKNGVLYDVSFALAVSNH 132
Qy      122 PLNNTPTVAGSPGLRELTSLTEILKGVLIORNPOLCQDPTILMKDIFHKNNOLAL 181
Db      133 -MNRQ-----GLRELPMKRLSEILLNGVKISNNPKLCMDVYLVNNDITDSRK-PL 182
Qy      182 TLID-TNNSRACHPCSPMCKSGRSGESSEDOQLSTRVYAGGCA-RCKGPLPTDCHEQ 239
Db      183 TVLIDFASNLSSCPKCHPNCETEDHCWAGEFQNCOTLTKVYCAQCGSGRCGVPSDCHNQ 242
Qy      240 CAAGCTGPKHSHCLACLFHNSGICELHCPALVYNTDTFESMPREGRYTTGACGVAC 299
Db      243 CAAGCTGPRESCICACRFRDATTCKDTCPPVLVYNPTTYOMDVPEGRYSFGATCVRBC 302
Qy      300 PYNVLTSDVGSCTVCPPLHNOEVTAEADGTQRCCKSPCAR 340
Db      303 PYNVYVTDHSGCVSRSCNTDTYEV-BENGVRCKCKCDGLCSK 342

RESULT 5
EGFR_HUMAN
ID EGFR_HUMAN STANDARD: PRT: 1210 AA.
AC P00533; P06268; Q14225;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112).
GN EGFR OR ERBB1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP MEDLINE=84219729; Pubmed=6328312;
RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
RA Mayer E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
RA "Human epidermal growth factor receptor cDNA sequence and aberrant
RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
RL Nature 309:418-425(1984).
RN [2]
RP SEQUENCE OF 713-924 FROM N.A.
RX MEDLINE=84196372; Pubmed=6326261;
RA Lin C.R., Chen W.S., Krueger W., Stolarsky L.S., Weber W.,
RA Evans R.M., Verma I.M., Gill G.N., Rosenfeld M.G.;
RT "Expression cloning of human EGF receptor complementary DNA: gene
RT amplification and three related messenger RNA products in A431
RT cells.";
RL Science 224:843-848(1984).
RN [3]
RP SEQUENCE OF 150-962 FROM N.A.
RX MEDLINE=84245835; Pubmed=6330563;
RA Xu Y., Ishii S., Clark A.D.L., Sullivan M., Wilson R.K., Ma D.P.,
RA Roe B.A., Merlino G.T., Pastan I.;
RT "Human epidermal growth factor receptor cDNA is homologous to a
RT variety of RNAs overproduced in A431 carcinoma cells.";
RL Nature 309:806-810(1984).
RN [4]
RP SEQUENCE OF 1028-1210 FROM N.A.
RX MEDLINE=85046483; Pubmed=6093780;
RA Simmen F.A., Gope M.L., Schultz T.Z., Wright D.A., Carpenter G.,
RA O'Malley B.W.;
RT "Isolation of an evolutionarily conserved epidermal growth factor
RT receptor cDNA from human A431 carcinoma cells.";
RL Biochem. Biophys. Res. Commun. 124:125-132(1984).
RN [5]
RP SEQUENCE OF 1-29 FROM N.A.

```

```

RX MEDLINE=88217333; Pubmed=3329716;
RA Haley J., Whittle N., Bennett P., Kinchington D., Ullrich A.,
RA Waterfield M.;
RT "The human EGF receptor gene: structure of the 110 kb locus and
RT identification of sequences regulating its transcription.";
RL Oncogene Res. 1:375-396(1987).
RN [6]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=91107677; Pubmed=1988448;
RA Haley J.D., Waterfield M.D.;
RT "Contributory effects of de novo transcription and premature
RT transcript termination in the regulation of human epidermal growth
RT factor receptor proto-oncogene RNA synthesis.";
RL J. Biol. Chem. 266:1746-1753(1991).
RN [7]
RP SEQUENCE OF 1-29 FROM N.A.
RX MEDLINE=85270438; Pubmed=2991899;
RA Ishii S., Xu Y.H., Strutton R.H., Roe B.A., Merlino G.T., Pastan I.;
RT "Characterization and sequence of the promoter region of the human
RT epidermal growth factor receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:4920-4924(1985).
RN [8]
RP SEQUENCE OF 540.
RA Kohda D.;
RL Submitted (SEP-1997) to the SWISS-PROT data bank.
RN [9]
RP RECEPTOR ACTIVITY.
RX MEDLINE=84191554; Pubmed=6325948;
RA Mroczkowski B., Mosig G., Cohen S.;
RT "ATP-stimulated interaction between epidermal growth factor receptor
RT and supercoiled DNA.";
RL Nature 309:270-273(1984).
RN [10]
RP PHOSPHORYLATION.
RX MEDLINE=89278137; Pubmed=2543678;
RA Margolis B.L., Lax I., Kris R., Dombalagian M., Honegger A.M.,
RA Hawk R., Givol D., Ullrich A., Schlessinger J.;
RT "All autophosphorylation sites of epidermal growth factor (EGF)
RT receptor and HER2/neu are located in their carboxyl-terminal tails.
RL Identification of a novel site in EGF receptor.";
RL J. Biol. Chem. 264:10667-10671(1989).
RN [11]
RP REVIEW.
RX MEDLINE=87297456; Pubmed=3039909;
RA Carpenter G.;
RT "Receptors for epidermal growth factor and other polypeptide
RT mitogens.";
RL Annu. Rev. Biochem. 56:881-914(1987).
RN [12]
RP FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
RX AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
RA VACCINIA VIRUS GROWTH FACTOR.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- MISCELLANEOUS: BINDING OF EGF TO THE RECEPTOR LEADS TO
CC INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
CC TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND
CC CELL PROLIFERATION.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X00588; CAA25240.1; -
DR EMBL: X06370; CAA29668.1; -
DR EMBL: X00663; CAA25282.1; -
DR EMBL: M38425; AAA63171.1; -
DR EMBL: M11234; AAA52370.1; -

```

DR PIR: A00641; GQHUE.
 DR PIR: A00642; GQHUE2.
 DR PIR: A23062; A23062.
 DR HSP: P11362; 1FG1.
 DR SWISS-2DPAGE; P00533; HUMAN.
 DR MIM: 131550; -.
 DR InterPro: IPR000494; EGFR.L.
 DR InterPro: IPR000719; Euk_pkinase.
 DR InterPro: IPR002174; Furin-like.
 DR InterPro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; Fu_3.
 DR SMART: SM00219; Tyrc; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 DR Transmembrane; Glycoprotein; Repeat; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 24
 FT CHAIN 25 1210
 FT DOMAIN 25 645
 FT TRANSMEM 646 668
 FT DOMAIN 669 1210
 FT REPEAT 75 300
 FT REPEAT 390 600
 FT REPEAT 1025 600
 FT DOMAIN 1025 1071
 FT DOMAIN 712 979
 FT NP_BIND 718 726
 FT BINDING 745 745
 FT ACT_SITE 837 837
 FT MOD_RES 678 678
 FT MOD_RES 1092 1092
 FT MOD_RES 1110 1110
 FT MOD_RES 1172 1172
 FT MOD_RES 1197 1197
 FT CARBOHYD 128 128
 FT CARBOHYD 175 175
 FT CARBOHYD 196 196
 FT CARBOHYD 332 332
 FT CARBOHYD 361 361
 FT CARBOHYD 413 413
 FT CARBOHYD 444 444
 FT CARBOHYD 528 528
 FT CARBOHYD 568 568
 FT CARBOHYD 603 603
 FT CARBOHYD 623 623
 FT CONFLICT 540 540
 FT SEQUENCE 1210 AA; 134277 MW; DBA2A50B4EFB6ED2 CRC64;
 Query Match 34.7%; Score 793; DB 1; Length 1210;
 Best Local Similarity 45.3%; Pred. No. 2.2e-54;
 Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

DB 244 RESDCLVCRKFRDEATCTDTCPEMLLYNPTTYQMKNVNEGKTSRGATCVKCKPRNYVTVD 303
 QY 308 VGSCTVLCPLHNOEVTAEDEGTQRCCKSCPKAR 340
 DB 304 HGSVCVACGADSYEM-EEDGVKCKCKCEGPCRK 335
 RESULT 6
 EGF_MOUSE
 ID EGF_MOUSE STANDARD; PRT; 1210 AA.
 AC Q01279;
 DT 01-FEB-1996 (Rel. 33, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (EC 2.7.1.112).
 GN EGF.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-Liver;
 RX MEDLINE-93026370; PubMed-1408137;
 RA Avivi A., Skorecki K., Yayon A., Givol D.;
 RT "Promoter region of the murine fibroblast growth factor receptor 2
 (bek/Kcfr) gene.";
 RL Oncogene 7:1957-1962(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; AND CD-1; TISSUE-Liver, and Decidua;
 RX MEDLINE-93126380; PubMed-7678348;
 RA Paria B.C., Das S.K., Andrews G.K., Dey S.K.;
 RT "Expression of the epidermal growth factor receptor gene is regulated
 in mouse blastocysts during delayed implantation.";
 RL Proc. Natl. Acad. Sci. U.S.A. 90:55-59(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C; TISSUE-Liver;
 RA Hildes M.L.;
 RT Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-B6/C3; TISSUE-Liver;
 RX MEDLINE-94170986; PubMed-8125255;
 RA Luetteke N.C., Phillips H.K., Qiu T.H., Copeland N.G., Barp H.S.,
 RA Jenkins N.A., Lee D.C.;
 RT "The mouse waved-2 phenotype results from a point mutation in the EGF
 receptor tyrosine kinase.";
 RL Genes Dev. 8:399-413(1994).
 RN [5]
 RP SEQUENCE OF 1-714 FROM N.A.
 RC TISSUE-Brain;
 RA Avivi A., Iax I., Ullrich A., Schlessinger J., Givol D., Morse B.;
 RT "Comparison of EGF receptor sequences as a guide to study the ligand
 binding site.";
 RL Oncogene 6:673-676(1991).
 RN [6]
 RP SEQUENCE OF 969-1117 FROM N.A.
 RC STRAIN-C3H;
 RA Estinger D.P., Serrero G.;
 RT Submitted (Jun-1992) to the EMBL/GenBank/DBJ databases.
 RN [7]
 RP FUNCTION: THE EGF RECEPTOR MEDIATES THE BIOLOGICAL SIGNAL OF EGF,
 AND ALSO OF TGF-ALPHA, AMPHIREGULIN, HEPARIN-BINDING EGF, GP30 AND
 VACCINIA VIRUS GROWTH FACTOR (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
 PROTEIN TYROSINE PHOSPHATE.
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -1- INTERACTION: BINDING OF EGF TO THE RECEPTOR LEADS TO
 INTERNALIZATION OF THE EGF-RECEPTOR COMPLEX, INDUCTION OF THE
 TYROSINE KINASE ACTIVITY, STIMULATION OF CELL DNA SYNTHESIS, AND

```

CC      CELL PROLIFERATION.
CC      -I- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isdb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
DR      EMBL; X78987; CAA55587.1; -
DR      EMBL; U03425; AAA17899.1; -
DR      EMBL; X59698; CAA42219.1; -
DR      EMBL; L06864; AAA53029.1; -
DR      HSSP; Z12608; CAA78249.1; -
DR      HSSP; P13362; IFC1.
DR      MGD; MGI:95294; Egfr.
DR      InterPro: IPRO00494; EGFR_L.
DR      InterPro: IPRO00719; Euk_Pkinase.
DR      InterPro: IPRO02174; Furin-like.
DR      InterPro: IPRO01245; Tyr_Kin.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF00069; pkinase; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      SMART; SM00261; FU; 3.
DR      SMART; SM00219; TyrcKc; 1.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.
DR      PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
KW      Transmembrane; Glycoprotein; Receptor; Signal; Transferase;
KM      Tyrosine-protein kinase; ATP-binding; Phosphorylation; Repeat.
FT      SIGNAL          1         24
FT      CHAIN           25        1210
FT      DOMAIN          25         647
FT      TRANSMEM       648         670
FT      DOMAIN          671        1210
FT      REPEAT          75          300
FT      REPEAT          390          600
FT      DOMAIN          1028        1071
FT      NP_BIND         714          981
FT      BINDING         747          747
FT      ACT_SITE        839          839
FT      MOD_RES         680          680
FT      MOD_RES         1092         1092
FT      MOD_RES         1110         1110
FT      MOD_RES         1172         1172
FT      MOD_RES         1197         1197
FT      MOD_RES         1197         1197
FT      CARBOHYD        128          128
FT      CARBOHYD        175          175
FT      CARBOHYD        196          196
FT      CARBOHYD        352          352
FT      CARBOHYD        413          413
FT      CARBOHYD        444          444
FT      CARBOHYD        528          528
FT      CARBOHYD        568          568
FT      CARBOHYD        603          603
FT      CARBOHYD        623          623
FT      CONFLICT        19            19
FT      CONFLICT        539          539
FT      CONFLICT        991          991
FT      CONFLICT        1116         1117
SO      SEQUENCE        1210 AA; 134853 MW; 690B20D46DF2DAF5 CRC64;

Query Match      34.5%, Score 789; DB 1; Length 1210;
Best Local Similarity 46.3%; Pred. No. 4.5e-54;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

```

Db	14	ELFALCAAGALAEKVKVCGQGSNRLPDGLTFEEDHFLISLQWYNNNCEVAVGNLEITFYQGN	73
Qy	69	ASLSFLDIDIEVQGYVLIANNOVROVPLQRLRYRGTOLEFEDNYALAVLDNGDPLNTPP	128
Db	74	YDLSEFLKTIQIEVAGYVLIANTVERPLEMLQIIRGNALIENTYVALADLSN-----	124
Qy	129	VTGASPGGLRELOARSLTELLKGSVLIQRNPOLCYODTLMKRI-----FHKNQGLATLLI	184
Db	125	-YGNTRGLRELPRNLQDELILGAVRESNPILCNDGTLQWRIDYONVPSNNSML- --	180
Qy	185	DTNNSRACHPSPKSCRSKRSRGESSEEDCSLFTPTVCAGCA-RCKGPLPTDCCHEQCAG	243
Db	181	-QSHPSGPCPKDPCSPRGSCWGGEEENCKQLTILICAGQCSHRCRGRSPSDCHNCAAG	239
Qy	244	CTGRKHSDCIACLFHFNHSGICELHCPALVTYNTDFESMPEKGRYTFGASCVTACPNY	303
Db	240	CTGRSPSCILVCKQFQDEATCKDKPCPLMYNPTTYQMDVNPFGKYSFATCVKCPRY	299
Qy	304	LSTDVSGSLTLCPLHNOEYVAEDGTQCEKSCRPCAR 340	
Db	300	VVTDHSGCVACRAGDPYEV-EEDEIRKCKKCDPCRK 335	
RESULT 7			
ERB4_HUMAN	STANDARD:	PRT; 1308 AA.	
AC	Q15303		
DT	15-DEC-1998 (Rel. 37, Created)		
DT	15-DEC-1998 (Rel. 37, Last sequence update)		
DT	20-AUG-2001 (Rel. 40, Last annotation update)		
DE	ERB4 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)		
DE	(P180ERB4) (TYROSINE KINASE-TYPE CELL SURFACE RECEPTOR HER4).		
GN	ERB4 OR HER4.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.		
OX	NCBI_TaxID=9606;		
RN	(1)		
RP	SEQUENCE FROM N.A. (ISOFORM JM-A).		
RC	TISSUE=Breast carcinoma;		
RX	MEDLINE=93189574; PubMed=8383326;		
RA	PLOWMAN G.D., CLOUSCOU J.-M., WHITNEY G.S., GREEN J.M., CARLTON G.W.		
RA	FOY L., NEUBAUER M.G., SHOYAB M.;		
RT	"Ligand-specific activation of HER4/p180erb4, a fourth member of the		
RL	epidermal growth factor receptor family.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 90:1746-1750(1993).		
RN	[2]		
RP	SEQUENCE FROM N.A. (ISOFORMS JM-A AND JM-B).		
RC	TISSUE=fetal brain;		
RX	MEDLINE=97476287; PubMed=9334263;		
RA	ELENIUS K., CORFAS G., PAUL S., CHOI C.J., RIO C., PLOWMAN G.D.,		
RA	KLAGSGRUM M.;		
RT	"A novel junctional membrane domain isoform of HER4/erb4. Isoform-specific		
RT	tissue distribution and differential processing in response to		
RT	phorbol ester";		
RL	J. Biol. Chem. 272:26761-26768(1997).		
CC	-1- FUNCTION: SPECIFICALLY BINDS AND IS ACTIVATED BY NEUREGULIN, NRG-		
CC	2. NRG-3, HEPARIN-BINDING EGF-LIKE GROWTH FACTOR, BETACELLULIN AND		
CC	NTAK. INTERACTION WITH THESE FACTORS INDUCES CELL DIFFERENTIATION		
CC	NOT ACTIVATED BY EGF, TGF-A, AND AMPHIREGULIN.		
CC	-1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +		
CC	PROTEIN TYROSINE PHOSPHATE.		
CC	-1- SUBUNIT: HOMODIMER OR HETERODIMER WITH EACH OF THE OTHER ERBB		
CC	RECEPTORS (POTENTIAL).		
CC	-1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.		
CC	-1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; JM-A (SHOWN HERE) AND JM-B;		
CC	ARE PRODUCED BY ALTERNATIVE SPLICING. THE TWO ISOFORMS DIFFER		
CC	FUNCTIONALLY IN THEIR RESPONSE TO PHORBOL ESTER: JM-A IS PROCESSED		
CC	BUT NOT JM-B. SO, THEY RESPECTIVELY REPRESENT CLEAVABLE AND		
CC	NONCLEAVABLE FORMS OF THE RECEPTOR. BOTH ISOFORMS ARE EXPRESSED IN		
CC	CEREBELLUM, BUT ONLY THE JM-B ISOFORM IS EXPRESSED IN THE HEART.		
CC	-1- TISSUE SPECIFICITY: EXPRESSED AT HIGHEST LEVELS IN BRAIN, HEART,		
CC	KIDNEY, IN ADDITION TO SKELETAL MUSCLE, PARATHYROID, CEREBELLUM,		

CC PITUITARY, SPLEEN, TESTIS AND BREAST. LOWER LEVELS IN THYMUS,
CC LUNG, SALIVARY GLAND, AND PANCREAS.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE
CC RESIDUES.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: L07868; AAB59446.1; -
DR MIM: 600543; -
DR InterPro: IPR000494; EGR_L.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 4.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
KW Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
KW Transferase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KW Alternative splicing.
KW SIGNAL 1 25
FT CHAIN 26 1308
FT DOMAIN 26 651
FT TRANSMEM 652 675
FT DOMAIN 676 1308
FT DOMAIN 186 334
FT DOMAIN 496 633
FT DOMAIN 713 988
FT NP_BIND 724 732
FT BINDING 751 751
FT ACT_SITE 843 843
FT MOD_RES 1162 1162
FT MOD_RES 1188 1188
FT MOD_RES 1258 1258
FT MOD_RES 1284 1284
FT CARBOHYD 138 138
FT CARBOHYD 174 174
FT CARBOHYD 181 181
FT CARBOHYD 253 253
FT CARBOHYD 358 358
FT CARBOHYD 410 410
FT CARBOHYD 473 473
FT CARBOHYD 495 495
FT CARBOHYD 548 548
FT CARBOHYD 576 576
FT CARBOHYD 620 620
FT VASAPLIC 648 648
FT SEQUENCE 1308 AA; 146807 MW; 5E4AE8095D8761 CRC64;
(IN ISOCOREM JM-B)
Query Match 33.9%; Score 775; DB 1; Length 1308;
Best Local Similarity 45.1%; Pred. No. 6.1e-53;
Matches 157; Conservative 45; Mismatches 124; Indels 22; Gaps 7;
QY 9 WGLTALLPPGAA-----STOVCTGDMKRLPASPETHLDMRLHYLCQGVVQGNLEETLY 64
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 8 WWWWLVAAGVYVPSDQSGVACGKENTLSSLDQGYRALRKRYKENCEVVMGLLEITS 67
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 65 LPTNASLSFLDIDQEVQVLIANQVQVPLQRLRIYRGTDLPEDNALAVLNDGDPIN 124
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 68 IEHNBDLSFLRSREVTVYVALNQPFYLLPLENRIITNGTLYEDRYALALFLNVRKDQ 127

QY 125 NTTPVTCASPGLRLAOLRSLEILKGVLIQRPOLCYOFTILMKDIFHKNNOLATLI 184
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 128 NF-----GLOELGKNTLEILNGVYVDQKFLCYATIMQDIVRMPWSNLTLY 178
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 185 DTNRSRACHPCSPMKSGRCNGESSEDCQSLTRVYAGGC-ARCKGLPLDCCHEQCAAG 243
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 179 STNGSGGCGRCHKSCGTG-RCWGPTEHNCQTLTRVYCAEQDCGRCYGVPVSDCCRRCCAGG 237
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 244 CTGPKHSDCLACLFHNSGICELHCPALVYNTDFESMPREGRYTFGASCYACAPNY 303
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 238 CSGPRDIDCFACMNPNDGACVYQCPOTFYVNPPTFQLEHFNPAKYTGACVYKCPHNF 297
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
QY 304 LSTDVSGCTLVCPLEHDEVTAEDEGTORCEKSCPCAR-----GTHSLT 346
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
DB 298 V-VDSGSCVRACPSKMEV-ENNGIKMKKPCDTDICPRACDGIIGSLM 343
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
RESULT 8
ERB3_HUMAN STANDARD; PRT; 1342 AA.
ID ERB3_HUMAN
AC P21860;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ERB3-3 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112)
DE (TYROSINE KINASE-TYPE CELL SURFACE RECEPTOR HER3).
GN ERB3 OR HER3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_TaxID:9606;
RN [1]
RP MEDLINE=90083234; PubMed=2687875;
RX MEDLINE=90083234; PubMed=2687875;
RA Krius M.H., Issing W., Miki T., Popescu N.C., Aaronsen S.A.;
RT "Isolation and characterization of ERB3, a third member of the
RT ERBB/epidermal growth factor receptor family: evidence for
RT overexpression in a subset of human mammary tumors";
RL Proc. Natl. Acad. Sci. U.S.A. 86:9193-9197(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=90311312; PubMed=2164210;
RA Plowman G.J.D., Whitley G.S., Neubauer M.G., Green J.M., McDonald V.L.,
RA Todaro G.J., Shoyab M.;
RT "Molecular cloning and expression of an additional epidermal growth
RT factor receptor-related gene";
RL Proc. Natl. Acad. Sci. U.S.A. 87:4905-4909(1990).
RN [3]
RP SEQUENCE FROM N.A. (SHORT FORM).
RX MEDLINE=93282822; PubMed=7685162;
RA Katoh M., Yazaki Y., Sugimura T., Terada M.;
RT "c-erbB3 gene encodes secreted as well as transmembrane receptor
RT Tyrosine kinase";
RL Biochem. Biophys. Res. Commun. 192:1189-1197(1993).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULIN AND NPAK.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG FORM) AND
CC SECRETED (SHORT FORM).
CC -1- ALTERNATIVE PRODUCTS: TWO FORMS, A LONG TRANSMEMBRANE FORM AND A
CC SHORT SECRETED FORM, EXIST DUE TO ALTERNATIVE SPLICING.
CC -1- TISSUE SPECIFICITY: EPITHELIAL TISSUES AND BRAIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE (BY SIMILARITY).
CC -1- DISEASE: OVEREXPRESSED IN A SUBSET OF HUMAN MAMMARY TUMORS.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.


```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M29366; AAA35790.1; -
DR EMBL: M34309; AAA35979.1; -
DR EMBL: S61953; AAB26935.1; -
DR PIR: A36223; A36223.
DR HSSP: P1362; 1FC1.
DR MIM: 190151; -
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L_domain; 2.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; TYRC; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; FALSE_NEG.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR Transmembrane; Glycoprotein; Multigene family; Receptor; Signal;
DR Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;
KM Alternative splicing.
FT SIGNAL 1 19 POTENTIAL.
FT CHAIN 20 1342 ERBB-3 RECEPTOR PROTEIN-TYROSINE KINASE.
FT DOMAIN 20 643 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 644 664 POTENTIAL.
FT DOMAIN 665 1342 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 709 966 PROTEIN KINASE.
FT NP_BIND 715 723 ATP (BY SIMILARITY).
FT BINDING 742 742 ATP (BY SIMILARITY).
FT ACT_SITE 834 834 BY SIMILARITY.
FT CARBOHYD 126 126 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 250 250 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 408 408 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 414 414 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 437 437 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 469 469 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 522 522 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 566 566 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 616 616 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 141 183 EILSGVYIEKNDKLCHEMDTIDMDRIYDRDAEIVKDNGR
SC -> GQPMVPSGLTFPOADMDYLDLDDDRLLTLSSSK
VPTVLAAY (IN SHORT ISOFORM).
FT VARSPLIC 184 1342 MISSING (IN SHORT ISOFORM).
FT CONFLICT 560 560 E -> G (IN REF. 2).
FT CONFLICT 1064 1064 E -> G (IN REF. 2).
SQ SEQUENCE 1342 AA; 148097 MW; 7201E7F66CA374BD CRC64;
Query Match 32.2%; Score 735.5; DB 1; Length 1342;
Best Local Similarity 44.0%; Pred. No. 7.7e-50;
Matches 150; Conservative 49; Mismatches 121; Indels 21; Gaps 9;

```

```

OY 188 RSRACHPSPMKSGRCWGESEDCOSLTRYACAGG--ARCKGLPTDCCHRCACGCTG 246
DB 179 NGRSCPCHCEVKG--RCWPGSEDCQTLTKITCAQCNGHCGPNPNOCCHECAGGSG 237
OY 247 PKHSCLACILHNHSGICELHCPALVTNTDTFFESMPNDEGRYTGASCVTPAPNYIST 306
DB 238 PDDTCCFACRHNHNDGACVPCRPOLVYNNKLTFOLEPRLPHKRYQYGVASCPHNFV-V 296
OY 307 DVGSCTLVCPLEHNDVTAEDGTQRCCKSPCAR--GTHS 344
DB 297 DQTSVACRPDPKMEVD--KNGLKCEPCGGLCPKACEGTGS 336
RESULT 9
ERB3_RAT STANDARD: PRT; 1339 AA.
AC Q62799; Q62955;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE ERBB-3 RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112).
GN ERBB3.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Liver;
RX MEDLINE=96096535; PubMed=8522190;
RA Hellyer N.J., Kim H.-H., Greaves C.H., Sterke S.L., Koland J.G.;
RT "Cloning of the rat ErbB3 cDNA and characterization of the
RT recombinant protein."
RL Gene 165:279-284(1995).
RN [2]
RP REVISION TO 85.
RA Hellyer N.J.;
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 922-1097 FROM N.A.
RC STRAIN-SPRAGUE-DAWLEY; TISSUE=Sciatic nerve;
RX MEDLINE=97184212; PubMed=9030624;
RA Carroll S.L., Miller M.L., Frohner P.W., Kim S.S., Corbett J.A.;
RT "Expression of neurogins and their putative receptors, ErbB2 and
RT ErbB3, is induced during Wallerian degeneration."
RL J. Neurosci. 17:1642-1659(1997).
CC -1- FUNCTION: BINDS AND IS ACTIVATED BY NEUREGULINS AND NTAK.
CC -1- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
CC PROTEIN TYROSINE PHOSPHATE.
CC -1- SUBUNIT: HETERODIMER WITH EACH OF THE OTHER ERBB RECEPTORS
CC (POTENTIAL).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
CC -1- DOMAIN: THE CYTOPLASMIC PART OF THE RECEPTOR MAY INTERACT WITH THE
CC SH2 OR SH3 DOMAINS OF MANY SIGNAL-TRANSDUCING PROTEINS.
CC -1- PTM: LIGAND-BINDING INCREASES PHOSPHORYLATION ON TYROSINE RESIDUES
CC AND PROMOTES ITS ASSOCIATION WITH THE P85 SUBUNIT OF
CC PHOSPHATIDYLINOSITOL 3-KINASE.
CC -1- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U29339; AAC28498.1; -
DR EMBL: U52530; AAC35050.1; -
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.

```

DR Interpro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep.L.domain; 2.
 DR PRINTS: PRO0109; TYRKINASE.
 DR SMART: SM00261; FU; 5.
 DR SMART: SM00219; TYRKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; FALSE_NEG.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane, Glycoprotein; Multigene family; Receptor; Signal;
 KM Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation.
 FT SIGNAL 1 19
 FT CHAIN 20 1339
 FT DOMAIN 20 643
 FT TRANSMEM 644 662
 FT DOMAIN 663 1339
 FT DOMAIN 707 964
 FT NP_BIND 713 721
 FT BINDING 740 740
 FT ACT_SITE 832 832
 FT CARBOHYD 126 126
 FT CARBOHYD 250 250
 FT CARBOHYD 353 353
 FT CARBOHYD 408 408
 FT CARBOHYD 414 414
 FT CARBOHYD 437 437
 FT CARBOHYD 469 469
 FT CARBOHYD 522 522
 FT CARBOHYD 566 566
 FT CARBOHYD 616 616
 FT CONFLICT 1028 1028
 FT SEQUENCE 1339 AA; 147619 MW; F120281D432423D8 CRC64;

Query Match 30.5%; Score 698; DB 1; Length 1339;
 Best Local Similarity 42.1%; Pred. No. 6,6e-47;
 Matches 147; Conservative 46; Mismatches 130; Indels 26; Gaps 10;

OY 3 LAAALCWGILLALPPGAA---STOYCTGTDMKRLRPAPEHMLRLHLYGCCVVOGN 59
 DB 7 LQVLC---FLSLIARGSEMGNSOAVCPGTUNGSLSTGADANOYOTLYLYKECEVMGN 62
 OY 60 LETTYLPTNALSFLDIOEVQGVYLAHNOVQVPLQRLRYRGTLQEDNYALVLN 119
 DB 63 LELVLGNHADSFLQWIMEVGYVLAHNEFSVPLPLRLRYRGVQVVDGFAIVM-- 120
 OY 120 GDLPLNTPTVGTGASPGGLRELRLSRLTEILKGVLIQRPQCLCYDTILMKDIFHKNOL 179
 DB 121 ---LNYNT---NSSHALQLKFTQLEILSGVYTEKNDKLCMDITIMRDIVRVR-- 170
 OY 180 ALTLIDTNSRACHPCSPCKGSRGSESESDOSLITFVAGGC-ARKGPIPTCCHE 238
 DB 171 GAEIVYKKNGANCPCHVEYCKG-RCWGPGRDDQILTKTICAPQCGRGFPNPOCCHD 229
 OY 239 OCAAGCTGPKHSKDLACLAFHNSGICELCPALVYNTPTFESMPNDEGRYFGASCYTA 298
 DB 230 ECAGCGSGGQDDPCFCACRFENDSGACVPRCPRELYVNNKTLFQLEPRPHKRYQGVGVAS 289
 OY 299 CPYNYLSTDVSGCTLVCPILHNOEVTAEDETQRCCKSPCAR---GTHS 344
 DB 290 CPHNEV-VDQTEFCVRACPPDKMEVD-KHGLKMCPCGGLCPACKEGTGS 336
 RESULT 10
 XMRK_XIPMA STANDARD: PRT; 1167 AA.
 AC P13388;
 DT 01-JAN-1990 (Rel. 13, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MELANOMA RECEPTOR PROTEIN-TYROSINE KINASE PRECURSOR (EC 2.7.1.112).
 GN XMRK OR TU.

OS Xiphophorus maculatus (Southern platyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
 OC Acanthomorpha; Acanthopterygii; Percomorpha; Atherinomorpha;
 OC Cyprinodontiformes; Poeciliidae; Xiphophorus.
 OX NCBI_TaxID=8083;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90015140; PubMed=2797166;
 RA Witbrodt J., Adam D., Maltischek B., Meueler W., Raulf F.,
 RA Telling A., Robertson S.M., Scharl M.,
 RT "Novel putative receptor tyrosine kinase encoded by the melanoma-
 RL inducing tu locus in xiphophorus.";
 RL Nature 341:415-421(1989).
 RP REVISION TO 515.
 RA Scharl M.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: PROBABLE RECEPTOR WITH PROTEIN-TYROSINE KINASE ACTIVITY.
 CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE - ADP +
 CC PROTEIN TYROSINE PHOSPHATE.
 CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
 CC -!- DISEASE: INVOLVED IN PIGMENT CELLS MALIGNANT MELANOMAS.
 CC -!- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X16891; CAA34770.2; -.
 CC PIR: S06142; S06142.
 DR HSSP: P1362; IEG1.
 DR Interpro: IPR000494; EGFR_L.
 DR Interpro: IPR000719; Euk_pkinase.
 DR Interpro: IPR002174; Furin-like.
 DR Interpro: IPR001245; Tyr_kin.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF00069; pkinase; 1.
 DR Pfam: PF01030; Recep.L.domain; 2.
 DR SMART: SM00261; FU; 3.
 DR SMART: SM00219; TYRKC; 1.
 DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
 DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
 DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
 KW Transmembrane, Glycoprotein; Receptor; Signal; Transferase;
 KW Tyrosine-protein kinase; ATP-binding; Phosphorylation; Proto-oncogene.
 FT SIGNAL 1 25
 FT CHAIN 26 1167
 FT DOMAIN 26 642
 FT TRANSMEM 643 665
 FT DOMAIN 666 1167
 FT DOMAIN 710 977
 FT NP_BIND 716 724
 FT BINDING 743 743
 FT ACT_SITE 835 835
 FT CARBOHYD 114 114
 FT CARBOHYD 144 144
 FT CARBOHYD 201 201
 FT CARBOHYD 356 356
 FT CARBOHYD 365 365
 FT CARBOHYD 398 398
 FT CARBOHYD 417 417
 FT CARBOHYD 501 501
 FT CARBOHYD 576 576
 FT CARBOHYD 621 621
 FT SEQUENCE 1167 AA; 129934 MW; 4793E0749DC1D55A CRC64;

RA	Thomas K. ;	to the EMBL/GenBank/DBJ databases.
RL	Submitted (MAR-1996)	
RA	(4)	
RP	MUTANTS.	
RX	MEDLINE=94147981; PubMed=8313880;	
RA	Arloian R.V., les G.M., Sternberg P.W.;	
RT	"Mutations in the Caenorhabditis elegans let-23 EGFR-like gene define	
RL	elements important for cell-type specificity and function.",	
CC	EMO J. 13:360-366(1994).	
CC	-i- FUNCTION: TYROSINE KINASE RECEPTOR REQUIRED FOR THE INDUCTION OF	
CC	C. ELEGANS VULVA. POSSIBLE RECEPTOR FOR THE INDUCTIVE SIGNAL	
CC	REQUIRED FOR VULVAL DEVELOPMENT. ACTIVATED BY LIN-3 AND ACTS BY	
CC	MAY OF LET-60 RAS. THE LIN-3/LET-23 PAIR IS A SIMPLIFIED VERSION	
CC	OF THE MAMMALIAN NEUREGULIN-ERBB NETWORK.	
CC	-i- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE -> ADP +	
CC	PROTEIN TYROSINE PHOSPHATE.	
CC	-i- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.	
CC	-i- SIMILARITY: BELONGS TO THE EGF RECEPTOR FAMILY.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/sib.ch).	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL: X57767; CAA40919.1; ALT SEQ.	
DR	EMBL: D63426; BAA09729.1; ALT_INT.	
DR	EMBL: Z70038; CAA93882.1; -.	
DR	PIR: S13422; S13422.	
DR	HSSP: P11362; IEG1.	
DR	Morpap: ZK1067.1; CE03840.	
DR	InterPro: IPR000494; EGFR_L.	
DR	InterPro: IPR000719; Euk_Dkinase.	
DR	InterPro: IPR002174; Rurin-like.	
DR	InterPro: IPR001245; Tyr_Kin.	
DR	Pfam: PF00757; Furin-like; 1.	
DR	Pfam: PF00069; pkinase; 1.	
DR	Pfam: PF01030; Recep_L_domain; 2.	
DR	SMART: SM00261; FU; 6.	
DR	SMART: SM00219; TyrKc; 1.	
DR	PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.	
DR	PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.	
KW	Transferrase; Tyrosine-protein kinase; ATP-binding; Phosphorylation;	
KW	Transmembrane; Glycoprotein; Receptor; Signal.	
FT	SIGNAL	1 22
FT	CHAIN	23 1323
FT	DOMAIN	23 818
FT	TRANSMEM	819 841
FT	DOMAIN	842 1323
FT	DOMAIN	885 1152
FT	NP_BIND	891 899
FT	BINDING	919 919
FT	ACT_SITE	1010 1010
FT	CARBOHYD	91 91
FT	CARBOHYD	169 169
FT	CARBOHYD	255 255
FT	CARBOHYD	376 376
FT	CARBOHYD	561 561
FT	CARBOHYD	655 655
FT	CARBOHYD	746 746
FT	CARBOHYD	776 776
FT	VARIANT	368 368
FT	VARIANT	469 469
FT	VARIANT	700 700
FT	VARIANT	753 753
FT	VARIANT	1065 1065
FT	VARIANT	1074 1074
QO	SEQUENCE	1323 AA: 150510 MW: 6B0307EE53EEFA99 CRC64;

```

Query Match      18.3%. Score 419; DB 1: Length 1323;
Best Local Similarity 29.0%; Pred. No. 4.3e-25;
Matches 102; Conservative 59; Mismatches 131; Indels 60; Gaps 14;

Oy    25 VCTGDMKLRTPASPEFHLDMLRHLYOCQCOVVOGNLELTYPNTN----- 68
       :|::|   |   |   |   |   |   |   |   |   |   |   |   |
Db    39 LCSGTNGISIKRGTONI-LEDLETMYRCGRARYGGALETITMANKIKRKRESTNSTYDPK 97
Oy    69 -----ASLSFLQDIIQEVGYVLIAHNOVRPLQRLIRVKTGLFEDNYALAVLNDGP 122
       :|::|   |   |   |   |   |   |   |   |   |   |   |   |
Db    98 NEDSPKSIINFEDNLIEIRGSILIVRANIQKISPRFAVIYGDEVEFDN-ALYIHKNDK- 155
Oy    123 LNNTPVTYGASBGIGREQLSLSTELTKGGVLIQNPNOLCY-QDTLKMDIFHNKNQIAL 181
       156 -----VHEVMRELRYIRNGSVTIQNPKMCIYGDKIIMDKELLYPD--VQ 199
Oy    182 TLIDNRSRACH-----PCSPMKGSGRCWGESSEDCOSLTRFYACAGCANC---KGPL 231
       :|   |   |   |   |   |   |   |   |   |   |   |   |   |
Db    200 KVEITNSHQHCYQNKSNMAKHCHESC-NKKCMSCGNDQRYRYSRPCSQCFFSNSSTS 258
Oy    232 PTDCHEOCAAGCTGPGPKHSDDLACLHFNHSGICELCFHALVYTNTDFESMPNPEGRTTF 291
       :||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db    259 SYECDSDACIGCGTGHGPKNCIACSKYEIDGICICTGCSRKIFFNKTRGLRVENPDGRYQN 318
Oy    292 GASCVTACPYNYL-STDVGSCTLYC-PLIHQDEVTAEDGTORCEKC-SKPQAR 340
       :|   |   |   |   |   |   |   |   |   |   |   |   |   |
Db    319 GNHCYKECPPELLIENDV--CVRHCSDDHHHYDATKD--VRECEKCRSSSCPK 366

RESULT_13
ID ILPR-BRALA STANDARD; PRT; 1363 AA.
AC 002466;
AD 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE INSULIN-LIKE PEPTIDE RECEPTOR PRECURSOR (EC 2.7.1.112) (ILP RECEPTOR).
OS Branchiostoma lanceolatum (Common lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
CC Branchiostoma.
CX NCBI_TaxId=7740;
[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=96408719; PubMed=8813726;
RX Pashmforoush M., Chan S.J., Steiner D.F.;
RT "Structure and expression of the insulin-like peptide receptor from amphioxus.";
RL ML
CC -!- FUNCTION: THIS RECEPTOR BINDS TO THE INSULIN RELATED PEPTIDE AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY.
CC -!- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP + PROTEIN TYROSINE PHOSPHATE.
CC -!- SUBUNIT: PROBABLE Tetramer OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE DOMAIN (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-PROTEIN KINASES.
CC -----
CC This SWISS-Prot entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation - CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/or send an email to license@isb-sib.ch).
CC -----
DR EMBL: S83394; AAB50848.1; .
DR HSPB; P06213; IIRK.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR001777; FN_IIL.

```

DR		InterPro:	IPR002174;	Furin-like.
DR		InterPro:	IPR002011;	Rcpotr_tyr_kin_II.
DR		InterPro:	IPR001245;	Tyr_kin.
DR	Pfam:	PF00041;	Fn3; 3.	
DR	Pfam:	PF00757;	Furin-Ilike; 1.	
DR	Pfam:	PF00069;	pkinase; 1.	
DR	Pfam:	PF01030;	Recep_L_domain; 1.	
DR	PRINTS:	PRO0014;	FMTPEIIT.	
DR	PRINTS:	PR00109;	TYRKINASE.	
DR	SMART:	SMD0060;	FN3; 3.	
DR	SMART:	SMD0281;	FU; 1.	
DR	SMART:	SMD0219;	TyrcKc; 1.	
DR	PROSITE:	PS00107;	PROTEIN_KINASE_AAP; 1.	
DR	PROSITE:	PS00109;	PROTEIN_KINASE_TYR; 1.	
DR	PROSITE:	PS00239;	RECEPTOR_TYR_KIN_II; 1.	
KW	Transferrase:	tyrosine-protein kinase: Receptor; Transmembrane: glycoprotein; ATP-binding; Phosphorylation; Signal. POTENTIAL.		
FT	SIGNAL	1 29		
FT	CHAIN	30	716	
FT				INSULIN-LIKE PEPTIDE RECEPTOR, ALPHA- CHAIN (POTENTIAL).
FT	PROPER	717	720	POTENTIAL.
FT	CHAIN	721	1363	INSULIN-LIKE PEPTIDE RECEPTOR, BETA- CHAIN (POTENTIAL).
FT				
FT	DMAIN	721	928	EXTRACELLULAR (POTENTIAL).
FT	TRANSMEM	929	949	POTENTIAL.
FT	DMAIN	950	1363	CYTOSOLASMIC (POTENTIAL).
FT	DMAIN	994	?	PROTEIN KINASE.
FT	NP_BIND	1000	1008	ATP (BY SIMILARITY).
FT	BINDING	1028	1028	AAP (BY SIMILARITY).
FT	ACT_SITE	1148	1148	BY SIMILARITY.
FT	MOT_RES	1174	1174	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	CARBOND	51	51	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	97	97	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	137	137	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	278	278	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	483	483	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	599	599	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	617	617	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	665	665	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	666	666	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	711	711	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	732	732	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	736	736	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	743	743	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	816	816	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	885	885	N-LINKED (GLCNAC...) (POTENTIAL).
FT	CARBOND	898	898	N-LINKED (GLCNAC...) (POTENTIAL).
SO	SEQUENCE	1363 AA;	154104 MW;	238120DBAENBIEID65 GRG64;

```

Query Match Similarity 15.0% Score 34.2-5; DB 1; Length 1363;
Best Local Similarity 28.3%; Pred. No.4,3e-19;
Matches 106; Conservative 40; Mismatches 124; Indels 105; Gaps 20.

OY 9 WGLL-----LALLPGCASTQVCTGTDMLRLPASPETHIDMLRHLYOGCQVQGNLELT 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 12 WAILPLIVIGLGLVPSNGEERYICDSMDINR-----VSNIRQL-ENCOTYIEGLQI- 61
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 64 YLPNLSISFLQDIDQEVQGVLIAHNQVQV---LQRLR-----IVRGTQ 106
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 62 -----LLIDFAEEQDYSGCLAFPNLVEITDYFLLYRVRGHTNSELFPNLVAVRGTN 112
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 107 LPEDNVATALVNDGPRLLNTTPYTGASPGGLRQLQRLSLTEIKGVLVLRQNPQLCYQDT 166
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 113 LF-FNALVVEFMLD-----MOKIGYSLQNTIRGSGVRLEKKNELCYDT 156
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 167 ILMKQDF---HNNQQLATLIDTNRGRAC-HPCSPMCK-----GSRCAESEDQCSLT 216
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 157 IDMSFLAEGYSNN---FIVDRREEECYVNFPCGCRIRKHPYLQDLCNAE--EHQKVC 210
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

OY 217 RTVCAGGCARCARKGLPTDCDCECAAGCTGPKHSDIACILHFNHSGICECLHCPALVTYNT 276
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

```

Db 211 PESOLGNR----DGISCCCHENIGGDDGPTERDCAACKFYVNBCLIQCPDPTQYK 266
Oy 277 D---TESMFPNPGSRV--TFGASCVYACAPNLYSTVGS---CTLVCPHLNQEVTAEDG 327
Db 267 DRRCTTECECCNTNTNSVWKLHHRKRCIECCPSGY--TTDINNRLCT----- 310
Oy 328 TORCE-KCSKPCARG 341
Db 311 --EEGCGCPKCKG 323

RESULT 14
ID INSR_DROME STANDARD: PRT: 2146 AA.
INSR_DROME
P09208: 024089: 024023:
01-MAR-1989 (Rel. 10, Created)
01-NOV-1997 (Rel. 35, Last sequence update)
30-MAY-2000 (Rel. 39, Last annotation update)
DE INSULIN-LIKE RECEPTOR PRECURSOR (EC 2.7.1.112).
GN INR OR INR-A OR DIR-A.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95354655; PubMed=7628438;
RA Fernandez R., Tabarini D., Azpiazu N., Frasch M., Schlessinger J.;
RT "The Drosophila insulin receptor homolog: a gene essential for
RT embryonic development encodes two receptor isoforms with different
RT signaling potential.";
RL EMBO J. 14:3373-3384(1995).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95181404; PubMed=7876183;
RA Ruan Y., Chen C., Cao Y., Garofalo R.S.;
RT "The Drosophila insulin receptor contains a novel carboxyl-terminal
RT extension likely to play an important role in signal transduction.";
RL J. Biol. Chem. 270:4236-4243(1995).
RN [3]
RP SEQUENCE OF 652-1749 FROM N.A.
RX STRAIN-OREGON-R; TISSUE=Embryo;
RC MEDLINE=87100165; PubMed=3099787;
RA Nishida Y., Hata W., Nishizuka Y., Rutter W.J., Ebina Y.;
RT "Cloning of a Drosophila cDNA encoding a polypeptide similar to the
RT human insulin receptor precursor.";
RL Biochem. Biophys. Res. Commun. 141:474-481(1986).
RN [4]
RP SEQUENCE OF 1297-1595 FROM N.A.
RX MEDLINE=86259667; PubMed=3014506;
RA Petruzzelli L., Herrera R., Arenas-Garcia R., Fernandez R.,
RA Birnbaum M.J., Rosen O.M.;
RT "Isolation of a Drosophila genomic sequence homologous to the kinase
RT domain of the human insulin receptor and detection of the
RT phosphorylated Drosophila receptor with an anti-peptide antibody.";
RL Proc. Natl. Acad. Sci. U.S.A. 83:4710-4714(1986).
RN [5]
RP FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN
RN AND HAS A TYROSINE-PROTEIN KINASE ACTIVITY (BY SIMILARITY).
RN [6]
RP CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +
RN PROTEIN TYROSINE PHOSPHATE.
RN [7]
RP SUBUNIT: TETRAMER OF 2 ALPHA AND 2 BETA CHAINS LINKED BY DISULFIDE
RN BONDS. THE ALPHA CHAINS CONTRIBUTE TO THE FORMATION OF THE LIGAND-
RN BINDING DOMAIN, WHILE THE BETA CHAIN CARRY THE KINASE DOMAIN.
RN [8]
RP SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN.
RN [9]
RP SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-
RN PROTEIN KINASES.
RN [10]
RN [11]
RN [12]
RN [13]
RN [14]
RN [15]
RN [16]
RN [17]
RN [18]
RN [19]
RN [20]
RN [21]
RN [22]
RN [23]
RN [24]
RN [25]
RN [26]
RN [27]
RN [28]
RN [29]
RN [30]
RN [31]
RN [32]
RN [33]
RN [34]
RN [35]
RN [36]
RN [37]
RN [38]
RN [39]
RN [40]
RN [41]
RN [42]
RN [43]
RN [44]
RN [45]
RN [46]
RN [47]
RN [48]
RN [49]
RN [50]
RN [51]
RN [52]
RN [53]
RN [54]
RN [55]
RN [56]
RN [57]
RN [58]
RN [59]
RN [60]
RN [61]
RN [62]
RN [63]
RN [64]
RN [65]
RN [66]
RN [67]
RN [68]
RN [69]
RN [70]
RN [71]
RN [72]
RN [73]
RN [74]
RN [75]
RN [76]
RN [77]
RN [78]
RN [79]
RN [80]
RN [81]
RN [82]
RN [83]
RN [84]
RN [85]
RN [86]
RN [87]
RN [88]
RN [89]
RN [90]
RN [91]
RN [92]
RN [93]
RN [94]
RN [95]
RN [96]
RN [97]
RN [98]
RN [99]
RN [100]
RN [101]
RN [102]
RN [103]
RN [104]
RN [105]
RN [106]
RN [107]
RN [108]
RN [109]
RN [110]
RN [111]
RN [112]
RN [113]
RN [114]
RN [115]
RN [116]
RN [117]
RN [118]
RN [119]
RN [120]
RN [121]
RN [122]
RN [123]
RN [124]
RN [125]
RN [126]
RN [127]
RN [128]
RN [129]
RN [130]
RN [131]
RN [132]
RN [133]
RN [134]
RN [135]
RN [136]
RN [137]
RN [138]
RN [139]
RN [140]
RN [141]
RN [142]
RN [143]
RN [144]
RN [145]
RN [146]
RN [147]
RN [148]
RN [149]
RN [150]
RN [151]
RN [152]
RN [153]
RN [154]
RN [155]
RN [156]
RN [157]
RN [158]
RN [159]
RN [160]
RN [161]
RN [162]
RN [163]
RN [164]
RN [165]
RN [166]
RN [167]
RN [168]
RN [169]
RN [170]
RN [171]
RN [172]
RN [173]
RN [174]
RN [175]
RN [176]
RN [177]
RN [178]
RN [179]
RN [180]
RN [181]
RN [182]
RN [183]
RN [184]
RN [185]
RN [186]
RN [187]
RN [188]
RN [189]
RN [190]
RN [191]
RN [192]
RN [193]
RN [194]
RN [195]
RN [196]
RN [197]
RN [198]
RN [199]
RN [200]
RN [201]
RN [202]
RN [203]
RN [204]
RN [205]
RN [206]
RN [207]
RN [208]
RN [209]
RN [210]
RN [211]
RN [212]
RN [213]
RN [214]
RN [215]
RN [216]
RN [217]
RN [218]
RN [219]
RN [220]
RN [221]
RN [222]
RN [223]
RN [224]
RN [225]
RN [226]
RN [227]
RN [228]
RN [229]
RN [230]
RN [231]
RN [232]
RN [233]
RN [234]
RN [235]
RN [236]
RN [237]
RN [238]
RN [239]
RN [240]
RN [241]
RN [242]
RN [243]
RN [244]
RN [245]
RN [246]
RN [247]
RN [248]
RN [249]
RN [250]
RN [251]
RN [252]
RN [253]
RN [254]
RN [255]
RN [256]
RN [257]
RN [258]
RN [259]
RN [260]
RN [261]
RN [262]
RN [263]
RN [264]
RN [265]
RN [266]
RN [267]
RN [268]
RN [269]
RN [270]
RN [271]
RN [272]
RN [273]
RN [274]
RN [275]
RN [276]
RN [277]
RN [278]
RN [279]
RN [280]
RN [281]
RN [282]
RN [283]
RN [284]
RN [285]
RN [286]
RN [287]
RN [288]
RN [289]
RN [290]
RN [291]
RN [292]
RN [293]
RN [294]
RN [295]
RN [296]
RN [297]
RN [298]
RN [299]
RN [300]
RN [301]
RN [302]
RN [303]
RN [304]
RN [305]
RN [306]
RN [307]
RN [308]
RN [309]
RN [310]
RN [311]
RN [312]
RN [313]
RN [314]
RN [315]
RN [316]
RN [317]
RN [318]
RN [319]
RN [320]
RN [321]
RN [322]
RN [323]
RN [324]
RN [325]
RN [326]
RN [327]
RN [328]
RN [329]
RN [330]
RN [331]
RN [332]
RN [333]
RN [334]
RN [335]
RN [336]
RN [337]
RN [338]
RN [339]
RN [340]
RN [341]
RN [342]
RN [343]
RN [344]
RN [345]
RN [346]
RN [347]
RN [348]
RN [349]
RN [350]
RN [351]
RN [352]
RN [353]
RN [354]
RN [355]
RN [356]
RN [357]
RN [358]
RN [359]
RN [360]
RN [361]
RN [362]
RN [363]
RN [364]
RN [365]
RN [366]
RN [367]
RN [368]
RN [369]
RN [370]
RN [371]
RN [372]
RN [373]
RN [374]
RN [375]
RN [376]
RN [377]
RN [378]
RN [379]
RN [380]
RN [381]
RN [382]
RN [383]
RN [384]
RN [385]
RN [386]
RN [387]
RN [388]
RN [389]
RN [390]
RN [391]
RN [392]
RN [393]
RN [394]
RN [395]
RN [396]
RN [397]
RN [398]
RN [399]
RN [400]
RN [401]
RN [402]
RN [403]
RN [404]
RN [405]
RN [406]
RN [407]
RN [408]
RN [409]
RN [410]
RN [411]
RN [412]
RN [413]
RN [414]
RN [415]
RN [416]
RN [417]
RN [418]
RN [419]
RN [420]
RN [421]
RN [422]
RN [423]
RN [424]
RN [425]
RN [426]
RN [427]
RN [428]
RN [429]
RN [430]
RN [431]
RN [432]
RN [433]
RN [434]
RN [435]
RN [436]
RN [437]
RN [438]
RN [
```


RN	[1]	SEQUENCE FROM N.A.
RP		
RA	Steele R.E., Mai N.H., Lieu P., Shenk M.A.;	
RL	Submitted (Apr-1991) to the EMBL/GenBank/DBJ databases.	
CC	-I- FUNCTION: THIS RECEPTOR PROBABLY BINDS AN INSULIN RELATED PROTEIN	
CC	AAD HAS A TYROSINE-PROTEIN KINASE ACTIVITY (BY SIMILARITY).	
CC	-I- CATALYTIC ACTIVITY: ATP + A PROTEIN TYROSINE = ADP +	
CC	PROTEIN TYROSINE PHOSPHATE.	
CC	-I- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).	
CC	-I- TISSUE SPECIFICITY: EXPRESSED IN DIVIDING EPITHELIAL CELLS.	
CC	-I- SIMILARITY: BELONGS TO THE INSULIN RECEPTOR FAMILY OF TYROSINE-	
CC	PROTEIN KINASES.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/	
CC	or send an email to license@isb-sib.ch).	
CC	-----	
DR	EMBL; M64612; AAA68205.1; .	
DR	HSSP; P06213; IIRK.	
DR	InterPro; IPR000494; EGFR_L.	
DR	InterPro; IPR000719; Erk_PKinase.	
DR	InterPro; IPR001777; FN_III.	
DR	InterPro; IPR002174; Furin-like.	
DR	InterPro; IPR002011; Receptor_tyr_kin_II.	
DR	InterPro; IPR001245; Tyr_kin.	
DR	Pfam; PF00041; Im3; 1.	
DR	Pfam; PF00757; Furin-like; 1.	
DR	Pfam; PF00069; pkinase; 1.	
DR	Pfam; PF01030; Recep_L_domain; 1.	
DR	PRINTS; PS00109; TYRKINASE.	
DR	SMART; SMO0060; FN3; 1.	
DR	SMART; SMO0261; FU; 1.	
DR	SMART; SMO0219; TYRKC; 1.	
DR	PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.	
DR	PROSITE; PS00109; RECEPTOR_TYR_KIN_II; 1.	
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.	
KM	Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;	
KW	Glycoprotein; Atg-binding; Phosphorylation; Signal.	
FT	SIGNAL	1..24
FT	CHAIN	25..1477
FT	DOMAIN	25..980
FT	TRANSMM	981..1001
FT	DOMAIN	1002..1477
FT	DOMAIN	1044..?
FT	NP_BIND	1050..1058
FT	BINDING	1077..1077
FT	ACT_SITE	1175..1175
FT	MOD_RES	1201..1201
FT	CARBOHYD	55..55
FT	CARBOHYD	255..255
FT	CARBOHYD	300..300
FT	CARBOHYD	325..325
FT	CARBOHYD	457..457
FT	CARBOHYD	491..491
FT	CARBOHYD	549..549
FT	CARBOHYD	644..644
FT	CARBOHYD	732..732
FT	CARBOHYD	791..791
FT	CARBOHYD	874..874
FT	CARBOHYD	895..895
FT	CARBOHYD	957..957
SO	SEQUENCE	1477 AA; 168276 MW; 74ACDBA7C56DE1D41 CRC64;

OY	51	OGCVOVGNL---	ELTLPYNASISFL	ODIOEYU	YLHANO	VOVPL--	ORLYIVGTO	106
Db	54	QNCSCWGNL	VKSTKYU	DEENFEP	RYPRKRL	REITGV	LLISLCP	LFKEPHLE
OY	107	LFEENYALAV	LNDGDP	PLNNTP	PTYGAS	PGGLRE	LOLSLE	ILKGV
Db	114	LIL--NYAL	VIYNN---	---	ELKEV	EPSP	LLI	NGGHN
OY	167	ILMKDI---	FHKNNOL	ALTLID	TNRSRA---	CHPC---	SPMG	SGSRWG---
Db	157	IRMSI	ILKDIHQ	TQGYIL-	ESNKL	MCDL	GLK	GHCHP
OY	207	--ESSE	DQSLTR	VYCSAG	SGCAR	CKGRPL	TPOCH	NOCAAG
Db	214	KQNKMA	QOQRE	CFNNQO-	GREGC	LDSD	HTIC	HNCSL
OY	262	GICELH	CPALTY	NTDTF--	ESMP-----	NPEGR	YTFG	ASCVTAC
Db	272	GQCVSK	CPR-KOYL	LVDF	FLCQ	SCP	YMSINST	EYHNH
OY	313	LVCPLH	NOEV	YTAED	GTOR	CEKC	334	
Db	324	-----	NNQ-----	TK	CEKC	333		

Search completed: April 11, 2002, 09:34:23
Job time: 310 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:35:06 ; Search time 40.45 Seconds
(without alignments)
1515.157 Million cell updates/sec

Title: US-09-234-208b-2
Perfect score: 2287
Sequence: 1 MELALCRWGLLALPPGA.....VGRGPPDAHVAVNLSRYEG 419

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :
1: SPREMBL_17:*
2: sp_archaea:*
3: sp_bacteria:*
4: sp_fungi:*
5: sp_human:*
6: sp_invertebrate:*
7: sp_mammal:*
8: sp_mhc:*
9: sp_organelle:*
10: sp_phage:*
11: sp_plant:*
12: sp_rodent:*
13: sp_virus:*
14: sp_vertebrate:*
15: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2275	99.5	419	09UK79	09UK79 homo sapien
2	1716.5	75.1	1259	6 018735	018735 canis famil
3	794	34.7	527	13 090836	090836 gallus gall
4	793	34.7	405	4 092795	092795 homo sapien
5	793	34.7	628	4 09H2C9	09H2C9 homo sapien
6	793	34.7	657	4 014226	014226 homo sapien
7	793	34.7	705	4 09G2X1	09G2X1 homo sapien
8	789	34.5	643	11 09ERV6	09ERV6 mus musculu
9	789	34.5	655	11 09WVF5	09WVF5 mus musculu
10	789	34.5	1210	11 09EP98	09EP98 mus musculu
11	781.5	34.2	1209	11 09QX70	09QX70 rattus norv
12	779.5	34.1	1308	11 09QZ27	09QZ27 rattus norv
13	773	33.8	473	11 09ESE0	09ESE0 rattus norv
14	734	32.1	331	4 09BUD7	09BUD7 homo sapien
15	723	31.6	149	6 09BG66	09BG66 oryctolagus
16	693	30.3	1165	13 09YH40	09YH40 xiphophorus
17	662.5	29.0	1328	13 P79754	P79754 tuagu rubrip
18	647	28.3	599	13 09PSH2	09PSH2 gallus gall
19	571.5	25.0	1433	5 09BIR9	09BIR9 anopheles g

20	463.5	20.3	1137	13 09W6F6	09W6F6 gallus gall
21	404.5	17.7	150	6 09BG64	09BG64 oryctolagus
22	402.5	17.6	1368	5 023821	023821 caenorhabdi
23	395	17.3	151	6 09BG65	09BG65 oryctolagus
24	363.5	15.9	366	5 026569	026569 schistosoma
25	363.5	15.9	1717	5 026566	026566 schistosoma
26	331	14.5	334	5 026567	026567 schistosoma
27	331	14.5	342	5 026568	026568 schistosoma
28	300	13.1	1472	5 09U5A8	09U5A8 bombyx mori
29	298.5	13.1	1671	5 09NUT5	09NUT5 biotomalari
30	298	13.0	1358	13 073798	073798 xenopus lae
31	298	13.0	1418	13 093457	093457 scophthalmu
32	294	12.9	2144	5 09VD94	09VD94 drosophila
33	290	12.7	1300	11 09QYH4	09QYH4 mus musculu
34	284	12.4	1362	13 09PVZ4	09PVZ4 xenopus lae
35	273.5	12.0	89	11 088459	088459 mus musculu
36	272	11.9	469	11 063721	063721 rattus norv
37	270	11.8	410	11 063720	063720 rattus norv
38	252.5	11.0	1371	11 09QYH4	09QYH4 mus musculu
39	232	10.1	1245	13 09YGH8	09YGH8 scophthalmu
40	230	10.1	946	5 09VJ04	09VJ04 drosophila
41	210	9.2	868	5 03VEE2	03VEE2 drosophila
42	203	8.9	1846	5 016131	016131 caenorhabdi
43	191	8.4	82	6 09NOK4	09NOK4 sus scrofa
44	189	8.3	131	5 09BH16	09BH16 anopheles g
45	181.5	7.9	91	11 088458	088458 mus musculu

ALIGNMENTS

RESULT 1

09UK79 ID 09UK79 PRELIMINARY; PRT; 419 AA.

AC 09UK79: 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)

DE HERSTATIN.

GN HER-2.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RX MEDLINE=99415951; PubMed=10485918;

RA Doherty J.K., Bond C., Jardim A., Adelman J.P., Clinton G.M.;

RT "The HER-2/neu receptor tyrosine kinase gene encodes a secreted

RT autolnhibitor.";

RL Proc. Natl. Acad. Sci. U.S.A. 96:10869-10874(1999).

RN [2]

RP SEQUENCE FROM N.A.

RA Doherty J.K., Clinton G.M., Adelman J.P., Evans A.J., Henner W.D.;

RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL: AF177761; AAD56009.2; -

DR InterPro: IPR000494; EGFRL.

DR InterPro: IPR002174; Furin-like.

DR Pfam: PF01030; Recep_L-like; 1.

DR SMART: SM00261; FU; 1.

SO SEQUENCE 419 AA; 45472 MW; FPCIBE347E2D030C CRC64;

Query Match 99.5%; Score 2275; DB 4; Length 419;

Best Local Similarity 99.5%; Pred. No. 4.9e-190;

Matches 417; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MELALCRWGLLALPPGAASVCTGTFDKLRLPASPEHMLRLKLYGCGVAGNL 60
DB 1 MELALCRWGLLALPPGAASVCTGTFDKLRLPASPEHMLRLKLYGCGVAGNL 60
QY 61 ELTVLPTNASLSFLDIOEVGVYLIANQVRQVPLRLRLIVRGTOFLFEDNYALAVLDNG 120

```
|||||
Db 61 EFTYLPANASLFDIDIOEGVGVLIANNOVAVPLQRLRIVRGQLTEDNVALALVING 120
OY 121 DPLNNTPTVTGASPGGLRELRLSLTEILKGVLIIQRPOLCYODTILMKDIFHKNNOLA 180
Db 121 DPLNNTPTVTGASPGGLRELRLSLTEILKGVLIIQRPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRIVCAGCARCKGRLPTDCCHQC 240
Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRIVCAGCARCKGRLPTDCCHQC 240
OY 241 AAGCGPRHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 AAGCGPRHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
OY 301 YNYLSTDVGSCITVCPPLNNOEVTAEADGTQRCCKSKPCARGTHSLPRPAVPVPLRMQP 360
Db 301 YNYLSTDVGSCITVCPPLNNOEVTAEADGTQRCCKSKPCARGTHSLPRPAVPVPLRMQP 360
OY 361 GFAHVPVLSFRLPSMDLVSAFYSLPLAPLSPVSPVSGVGRPDPAHVAVLSRYEG 419
Db 361 GFAHVPVLSFRLPSMDLVSAFYSLPLAPLSPVSPVSGVGRPDPAHVAVLSRYEG 419

RESULT 2
O18735 PRELIMINARY: PRT: 1259 AA.
AC 018735:
DT 01-JAN-1998 (TREMBLrel. 05, Created)
DT 01-JAN-1998 (TREMBLrel. 05, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE ERBB-2.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RA Yokota H.;
RL Submitted (OCT-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: AB008451; BAA23127.1; -.
DR HSSP: P06213; 1IRK.
DR InterPro: IPR002048; EF-hand.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_pkinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.
DR Pfam: PF01030; Recep_L-domain; 2.
DR SMART: SM00261; FU; 3.
DR SMART: SM00219; Tyrc; 1.
DR PROSITE: PS00018; EF_HAND; UNKNOWN_1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS00011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00109; PROTEIN_KINASE_TYR; 1.
KW ATP-binding; Transferase; Tyrosine-protein kinase.
SEQUENCE 1259 AA; 137989 MW; E37364D49C4ACD46 CRC64;
```

```
|||||
Db 121 DPLSGTAPAPAGGRLRELRLSLTEILKGVLIIQRPOLCHQDTILMKDIFHKNNOLA 180
OY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRIVCAGCARCKGRLPTDCCHQC 240
Db 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRIVCAGCARCKGRLPTDCCHQC 240
OY 241 AAGCGPRHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
Db 241 AAGCGPRHSDCLACLFHNHSGICELHCPALVTYNTDFESMPNDEGRYTFGASCVTACP 300
OY 301 YNYLSTDVGSCITVCPPLNNOEVTAEADGTQRCCKSKPCARCYGLGMHLREYAVASAN 360
Db 301 YNYLSTDVGSCITVCPPLNNOEVTAEADGTQRCCKSKPCARCYGLGMHLREYAVASAN 360
OY 356 LRMOPG--PAHVPVLSFRLPSMDLVSAFYSLPLAPLSPVSPVSGVGRPDPAHVAVLSRYEG 419
Db 356 LRMOPG--PAHVPVLSFRLPSMDLVSAFYSLPLAPLSPVSPVSGVGRPDPAHVAVLSRYEG 419

RESULT 3
O90836 PRELIMINARY: PRT: 527 AA.
AC 090836:
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE EGF/TGF-ALPHA RECEPTOR PRECURSOR.
GN C-ERBB.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92123214; PubMed=1732751;
RA Flickinger T.W., Maible N.J., Kung H.J.;
RT "An alternatively processed mRNA from the avian c-erbB gene encodes a
RT soluble, truncated form of the receptor that can block ligand-
RT dependent transformation."
RL Mol. Cell. Biol. 12:883-893(1992).
DR EMBL: M7637; AAA48759.1; -.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L-domain; 2.
DR SMART: SM00261; FU; 2.
KW Signal.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 527 EGF/TGF-ALPHA RECEPTOR.
SEQUENCE 527 AA; 58353 MW; 764564ABCC05258 CRC64;
```

```
Query Match 34.7%; Score 794; DB 13; Length 527;
Best Local Similarity 46.0%; Pred. No. 5,76-61;
Matches 155; Conservative 55; Mismatches 109; Indels 18; Gaps 7;
```

OY 244 CTPGPHSCCLACLFHNHSGIEIHELCPALTYNNDTFESMNPGRARYFGSCVTACAPYNY 303
 DB 250 CTGPHSCCLACRFRFDATCKKCPPLVLYNPTTYMDVNPBGKISFGATCYRECPHNT 309
 OY 304 LSTDVSCCTVCPPLHNOETVAEDGTQRCCKSCSKPCAR 340
 DB 310 VVFDHSGSCVRSCNFTDYEY-EENGVRCRCKKCDGLCSK 345
 RESULT 4
 092795 PRELIMINARY; PRT; 405 AA.
 AC 092795; 000732; 000688;
 DT 01-FEB-1997 (TREMBLrel. 02, Created)
 DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
 DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (P60 EPIDERMAL GROWTH
 DE FACTOR RECEPTOR).
 GN EGR.
 OS Homo sapiens (Human).
 OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 OX NCBI_TaxID=9606;
 RN 11
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=9078666; PubMed=8918811;
 RA Reiter J.L., Mathie N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN 12
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=95382957; PubMed=7654368;
 RA Ilekis J.V., Stark B.C., Scoccia B.;
 RT "Possible role of variant RNA transcripts in the regulation of
 RT epidermal growth factor receptor expression in human placenta.";
 RL Mol. Reprod. Dev. 41:149-156(1995).
 RN 13
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=64219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN 16
 RP SEQUENCE FROM N.A.
 RX MEDLINE=65267689; PubMed=2991749;
 RA Marinho G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Willson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.;
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.";
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 RN 17
 RP SEQUENCE FROM N.A.
 RA Reiter J.L., Thredgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Schein Sincclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,

```

RA Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA Maille N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative EGR transcripts encoding truncated receptor
RT isoforms";
RL Genomics 0:0-0(2000).
RN [8]
RP SEQUENCE FROM N.A.
RA Reiter J.L., Eley G.D.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; U48725; AAC50797.1; -
DR EMBL; U48722; AAC50802.1; -
DR EMBL; U48724; AAC50796.1; -
DR EMBL; U48726; AAC50798.1; -
DR EMBL; U48723; AAC50803.1; -
DR EMBL; U95089; AAB53063.1; -
DR EMBL; AF288738; AAC35787.1; -
DR InterPro: IPR000494; EGRF.L.
DR InterPro: IPR002174; Furin-like.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep_L-domain; 1.
DR SMART; SM00261; FU; 1.
KW Signal; Receptor.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 405 EPIDERMAL GROWTH FACTOR RECEPTOR.
SO SEQUENCE 405 AA; 44664 MW; F5DEB31787EF1822 CRC64;

Query Match 34.7%; Score 793; DB 4; Length 405;
Best Local Similarity 45.3%; Pred. No. 5,1e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4

QY 11 LLLALLPPGAA--STOVCTGTDMLRLPASPETHDMLRLHLYOGCGVVOGNELETYLPYN 68
   ||||| | :||| ||| :| | :| :| :||| ||| :|
DB 14 LLALCPASRALEKKKVCQGTSNKLTQGTGFEDHFLSLQRMFNCCVVLGNLEITYQGN 73
   ||||| | :||| ||| :| | :| :| :||| ||| :|

QY 69 ASLSFLDIOIEVQGVYLLAHNOVROPRLRIYKSTOLFEDNYTALAVLDNGDPLNNTPR 128
   ||||| :||| ||| ||| ||| :||| :| :||| ||| :|
DB 74 YDLSEFLKTYEAGVAGVLLALNTVERIPLENLOIIRGNMYEKSALAVLSNYD----- 126
   ||||| :||| ||| ||| ||| :||| :| :||| ||| :|

QY 129 VTGASPGGLREVLQSLPEFLKGVLIQNPOLCYQODTILMKDIFKKNQGLALTITDNR 188
   | :| :||| :||| ||| ||| ||| ||| :||| :| :||| ||| :|
DB 127 ---AKTKGTKELEPMNNLOELHGAVRFSNNPCLAVESIQMDVIYSSDLSLNNMSDFONH 183
   ||||| :||| ||| ||| ||| ||| ||| :||| :| :||| ||| :|

QY 189 SRACPCSPMCKSGKRCWGESSEDCSLRTFYAGGCA-RCKGPLPTDCHEQCAAGCTGP 247
   :| :| :| :| ||| ||| :||| ||| :| :| :| :| :||| ||| ||| |||
DB 184 LGSCKKCPSCPCNGSCWAGENECKLTKIICAQCSGRCRKSRSDDCHNCCMACTGP 243
   ||||| :||| ||| ||| ||| :||| :| :||| ||| :|

QY 248 KHSDDLACLFHNHSGICELCHPALVYNTYTFESMPNEGRYTFGASCVTACPYNYLSTD 307
   :| :| :| :| | :| :| :| :| :| :| :| :| :| :| :| :| :| :|
DB 244 RESDLCVORKFDEATCKDTCPLMLNPTTYQMDVNPPEGKYSFGATCYKKCRPNYVVD 303
   ||||| :||| ||| ||| ||| :||| :| :||| ||| :|

QY 308 VGSCTLVCPRLHNOEYTAEDGTORCKCKSPKAR 340
   ||||| :||| ||| ||| ||| :||| :| :||| ||| :|
DB 304 HGSCVRACGADSYEM-EDDGVKCKCKCEGPKCR 335
   ||||| :||| ||| ||| ||| :||| :| :||| ||| :|

RESULT 5
Q9H2C9 PRELIMINARY; PRT; 628 AA.
Q9H2C9
AC Q9H2C9;
DT 01-MAR-2001 (TREMBLrel. 16; Created)
DT 01-MAR-2001 (TREMBLrel. 16; Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17; Last annotation update)
DE TRUNCATED EPIDERMAL GROWTH FACTOR RECEPTOR.
GN EGRF.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=84219729; PubMed=6328312;

```

RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.:
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;
 RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.:
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.";
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Malhale N.J.:
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [4]
 RP SEQUENCE FROM N.A.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Scheel Sinclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balsubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Malhale N.J.:
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 0:0-0(2000).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Reiter J.L., Eley G.D.:
 RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF288738; AAG35790.1;
 DR InterPro: IPR000345; CycL_heme_bind.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 4.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 KW Receptor.
 SO SEQUENCE 628 AA; 69228 MW; 3A00A5511A3B6AE2 CRC64;

Query Match 34.7%; Score 793; DB 4; Length 628;
 Best local similarity 45.3%; Pred. No. 8,5e-61;
 Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;
 Oy 11 LLLALLPPGAA--STOYCTGDMKRLIPASPEPHLDMLRLYCGCOYVGNLELTPLTN 68
 Db 14 LLLALCPASALEEKRYCOGTSTNKLTLQGFEDHFLSTQFMFNNEVCEVLTITVORN 73
 Oy 69 ASLSPLDIDIOEGYVLIANOVROYPLORLIVRGTOJPEEDYVALVDNGPLNNTTP 128
 Db 74 YDLSFLKTIDEVAGYVLIANTVERIPLENLQIRGNMYENSTALAVLSNYD----- 126
 Oy 129 VTGASPGGLREJLRLSTELTLKGVLIQRPOLCYOTILMRDIFHKNQDLATLIDTNR 188
 Db 127 ---ANKTGKELPMRNQELILHGAVRSNNPALCNVESIMRWIVSSDFLSNNMFOQNH 183
 Oy 189 SRACHPCSPCKSGRCSESSDQCSTRTVCAGGCA-RCKGPLPLDCCHEGCAAGCTGP 247
 Db 184 LGSCKQKDDPCSPGSCGAGAEENQKLTFLITCAQCSGRCGRKSPDCHNQCAGCTGP 243
 Oy 248 KHSDDCLACHFNNSIGCELHCPALMYNTDTPFSMPREERYFGASCYACAYNTLSTD 307
 Db 244 RESDCLVCRFRDEATNCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCVKCRPNVYTD 303

Oy 308 VGSCTVCPHNOEYTAEDGTORCEKSCSKPCAR 340
 Db 304 HGSVRAAGCADSYEM-EDDGVKCKCKGCPCKR 335
 RESULT 6
 ID 014226 PRELIMINARY; PRT; 657 AA.
 AC 014226; 093010;
 DT 01-NOV-1996 (TREMBLrel. 01, Created)
 DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
 DE 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
 DE EPIDERMAL GROWTH FACTOR RECEPTOR (A431-SPECIFIC P115 EPIDERMAL GROWTH
 DE FACTOR RECEPTOR).
 GN EGFR.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;
 RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.:
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.";
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 RN [2]
 RP SEQUENCE OF 501-544 FROM N.A.
 RC TISSUE=PLACENTA;
 RX MEDLINE=97078686; PubMed=8918811;
 RA Reiter J.L., Malhale N.J.:
 RT "A 1.8 kb alternative transcript from the human epidermal growth
 RT factor receptor gene encodes a truncated form of the receptor.";
 RL Nucleic Acids Res. 24:4050-4056(1996).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ullrich A., Coussens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W.,
 RA Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J.,
 RA Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.:
 RT "Human epidermal growth factor receptor cDNA sequence and aberrant
 RT expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RL Nature 309:418-425(1984).
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;
 RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H.,
 RA Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A.,
 RA Hunts J.H., Shimizu N., Pastan I.:
 RT "Structure and localization of genes encoding aberrant and normal
 RT epidermal growth factor receptor RNAs from A431 human carcinoma
 RT cells.";
 RL Mol. Cell. Biol. 5:1722-1734(1985).
 RN [5]
 RP SEQUENCE FROM N.A.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J.,
 RA Scheel Sinclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L.,
 RA Balsubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
 RA Malhale N.J.:
 RT "Comparative genomic sequence analysis and isolation of human and
 RT mouse alternative EGFR transcripts encoding truncated receptor
 RT isoforms.";
 RL Genomics 0:0-0(2000).
 RN [6]
 RP SEQUENCE FROM N.A.
 RA Reiter J.L., Eley G.D.:
 RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; K03193; AAA52371.1;
 DR EMBL; U48728; AAC50800.1;
 DR EMBL; AF288738; AAG35788.1;

DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 3.
 KM Receptor.
 SQ SEQUENCE 657 AA; 72395 MW; CEC1D6C7C4C52C2A CRC64;

Query Match 34.7%; Score 793; DB 4; Length 657;
 Best Local Similarity 45.3%; Pred. No. 8.9e-61;

Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

QY 11 LLLALLPGAA--STOVCTGTMKRLPASPETHLMDLRHLYOGGQVYOGNLELYPTN 68
 DB 14 LLAALCPASRALEEKVKVCGTSNKLTQLGTFFEDHFLSLDRMNCCEVYVGNLEITYVORN 73
 QY 69 ASLSFLQDIQEVGYVLAHNOVROVPLQRLIRVGTQLFEDNYALAVLNDGDLNNTTP 128
 DB 74 YDLSEFKTIOEVAGYVLLALNTERIPLENLQIRGNMYENSVALAVLSND----- 126
 QY 129 VVGASPGRLRELQSLSTFIILKGVLIQNPOLCYQDTILMKDIFRKNQALATLIDTR 188
 DB 127 ---ANKTGILKELPMRNLQIILGAVRFSNNPALCNVESIQMRDIYSSDPLSNMSMDPFMH 183
 QY 189 SRACHPCSPMGKSGRCWGESSEDCSLFRVCAGGA-RCKGPLPTDCHEOCAGCTGP 247
 DB 184 LGSCKKCPSCPCNGSCWGEENCKLTKITICAOQSGRCRKSSDCHNOCAAGCTGP 243
 QY 248 KHSDELACLFHNHSGICELCPALVYNTDFESMPNDEGRYTFGASCVTACPYNYLSTD 307
 DB 244 RESDCLVCRKFEDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCVKCKPRNVVTD 303
 QY 308 VGSCTLVCPRLNQEYTAEDGTORCEKCSPPCAR 340
 DB 304 HGSCVRACGADSYEM-EEGVRKCKKCEGPCRK 335

RESULT 7
 Q9GZXI PRELIMINARY; PRT; 705 AA.
 AC Q9GZXI.
 DT 01-MAR-2001 (TREMblrel. 16, Created)
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)
 DE TRUNCATED EPIDERMAL GROWTH FACTOR RECEPTOR PRECURSOR (P110 EPIDERMAL GROWTH FACTOR RECEPTOR).
 DE EGFR.
 GN Homo sapiens (Human).
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OC NCBI_TaxID=9606;
 RX [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=PLACENTA;
 RA Reiter J.L., Threadgill D.W., Danielson A.J., Schehl C.M., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., Maibach N.J.;
 RA "Human and mouse alternative EGFR transcripts encoding only the extracellular domain of the receptor.";
 RT Submitted (Feb-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=84219729; PubMed=6328312;
 RA Ulrich A., Cousens L., Hayflick J.S., Dull T.J., Gray A., Tam A.W., Lee J., Yarden Y., Libermann T.A., Schlessinger J., Downward J., Mayes E.L.V., Whittle N., Waterfield M.D., Seeburg P.H.;
 RA "Human epidermal growth factor receptor cDNA sequence and aberrant expression of the amplified gene in A431 epidermoid carcinoma cells.";
 RT Nature 309:418-425(1984).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=85267689; PubMed=2991749;

RA Merlino G.T., Ishii S., Whang-Peng J., Knutsen T., Xu Y.-H., Clark A.J.L., Stratton R.H., Wilson R.K., Ma D.P., Roe B.A., Hunts J.H., Shimizu N., Pastan I.;
 RA "Structure and localization of genes encoding aberrant and normal epidermal growth factor receptor RNAs from A431 human carcinoma cells.";
 RT Mol. Cell. Biol. 5:1722-1734(1985).
 RN [4]

RP SEQUENCE FROM N.A.
 RX MEDLINE=97078686; PubMed=8918611;
 RA Reiter J.L., Maibach N.J.;
 RT "A 1.8 kb alternative transcript from the human epidermal growth factor receptor gene encodes a truncated form of the receptor.";
 RT Nucleic Acids Res. 24:4050-4056(1996).
 RN [5]

RP SEQUENCE FROM N.A.
 RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielson A.J., Schehl Sinclair C., Pearall R.S., Green P.J., Yee D., Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D., Maibach N.J.;
 RA "Comparative genomic sequence analysis and isolation of human and mouse alternative EGFR transcripts encoding truncated receptor isoforms.";
 RT Genomics 0:0-0(2000).
 RN [6]

RP SEQUENCE FROM N.A.
 RA Reiter J.L., Eley G.D.;
 RT Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.
 RL EMBL: AF15253; AAC3240.1;
 DR EMBL: AF288738; AAC35786.1;
 DR InterPro: IPR000345; CYC_heme_bind.
 DR InterPro: IPR000494; EGFR_L.
 DR InterPro: IPR002174; Furin-like.
 DR Pfam: PF00757; Furin-like; 1.
 DR Pfam: PF01030; Recep_L_domain; 2.
 DR SMART: SM00261; FU; 4.
 DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
 DR Signal; Receptor.
 FT SIGNAL 1 24
 FT CHAIN 25 705 POTENTIAL.
 FT TRUNCATED EPIDERMAL GROWTH FACTOR RECEPTOR.
 FT SEQUENCE 705 AA; 77312 MW; ACEF14942FF1650C CRC64;

Query Match 34.7%; Score 793; DB 4; Length 705;
 Best Local Similarity 45.3%; Pred. No. 9.7e-61;

Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

QY 11 LLLALLPGAA--STOVCTGTMKRLPASPETHLMDLRHLYOGGQVYOGNLELYPTN 68
 DB 14 LLAALCPASRALEEKVKVCGTSNKLTQLGTFFEDHFLSLDRMNCCEVYVGNLEITYVORN 73
 QY 69 ASLSFLQDIQEVGYVLAHNOVROVPLQRLIRVGTQLFEDNYALAVLNDGDLNNTTP 128
 DB 74 YDLSEFKTIOEVAGYVLLALNTERIPLENLQIRGNMYENSVALAVLSND----- 126
 QY 129 VVGASPGRLRELQSLSTFIILKGVLIQNPOLCYQDTILMKDIFRKNQALATLIDTR 188
 DB 127 ---ANKTGILKELPMRNLQIILGAVRFSNNPALCNVESIQMRDIYSSDPLSNMSMDPFMH 183
 QY 189 SRACHPCSPMGKSGRCWGESSEDCSLFRVCAGGA-RCKGPLPTDCHEOCAGCTGP 247
 DB 184 LGSCKKCPSCPCNGSCWGEENCKLTKITICAOQSGRCRKSSDCHNOCAAGCTGP 243
 QY 248 KHSDELACLFHNHSGICELCPALVYNTDFESMPNDEGRYTFGASCVTACPYNYLSTD 307
 DB 244 RESDCLVCRKFEDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCVKCKPRNVVTD 303
 QY 308 VGSCTLVCPRLNQEYTAEDGTORCEKCSPPCAR 340
 DB 304 HGSCVRACGADSYEM-EEGVRKCKKCEGPCRK 335

```

RESULT      8
O9ERV6      PRELIMINARY;      PRT:      643 AA.
AC      O9ERV6;
DT      01-MAR-2001 (TREMBlrel. 16, Created)
DT      01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT      01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE      EPIDERMAL GROWTH FACTOR RECEPTOR ISOFORM 2.
GN      EGFR.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N. A.
RC      STRAIN=C3H/101, 129/SVJ, 129/SVEV/TAC;
RA      Reiter J.L., Theadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA      Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA      Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA      Maible N.J.;
RT      *Comparative genomic sequence analysis and isolation of human and
RT      mouse alternative Egfr transcripts encoding truncated receptor
RT      isoforms.
RL      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AF275366; AAG28046.1; -.
DR      EMBL; AF275364; AAG28046.1; JOINED.
DR      EMBL; AF275365; AAG28046.1; JOINED.
DR      MGI; 95294; Egfr.
DR      InterPro; IPR000345; Cytc_heme_bind.
DR      InterPro; IPR000494; EGFR_L.
DR      InterPro; IPR002174; Furin-like.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      SMART; SM00261; FU; 4.
DR      PROSITE; PS00190; CYTOCHROME_C; UNKNOWN_1.
KW      Receptor.
SQ      SEQUENCE      643 AA; 71477 MW; DEF22002C84911B1 CRC64;

Query Match      34.5%; Score 789; DB 11; Length 643;
Best Local Similarity 46.3%; Pred. No. 1,9e-60;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

OY      11 LLLALLPGAA--STQVCTGTMDKLRLPASPETHLDMLRHLYGCCQVQGNLELYLPTN 68
DB      14 LLLALCAAGGALLEKKVCGGTSNRLTQLGTFEDHFLSLQRMVNNCEVVLGNLEIYVORN 73
OY      69 ASLSEFLQDIQEVQVYLIAHNOVROPRLQRLIAVGTQLEFDNYALAVLDNDGPLANTTP 128
DB      74 YDLSEFLKTIQEVAGVYLLAINTVERIPLENIQIRGNALYENTYALALISN----- 124
OY      129 VNGASGGGLREQLRLTEILKGVLIQRNPQLCYQDTILMKDI---FHKNNQLALTLI 184
DB      125 -YGTNNTGRLRELPRLMRLQELIGAVFNSNNPILCNMDITQMDIYQNVMSNMSML----- 180
OY      185 DTRNSRACHPCSPCKSGKSCRWGSESSDCSLRTVYAGGCA-RCKGPLPTDCCHEQCAAG 243
DB      181 -QSHPSCKPCPSKSCNGSCWGEGENCKLTKIICAAQCSHRCGRSPSCCHNQCAAG 239
OY      244 CTGPKHSDCLACILHNHSGICELHCPALVYTNDFESHPNPGRTTEGASCTYACPYV 303
DB      240 CTGPRSDCLVYCKQFQDEATCKDTCPLMLYNFTYQMDVNPBGKYSFGATCVKCPRYN 299
OY      304 LSTDVSGCTLVCPILHNOEYTAEDGTORCEKCSRPCAR 340
DB      300 VYTDGSCVACGPDYEV-EEEDGIRKCKKCGDPCRK 335

```

```

DT      01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT      01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE      EPIDERMAL GROWTH FACTOR RECEPTOR (EPIDERMAL GROWTH FACTOR RECEPTOR
DE      ISOFORM 3).
GN      EGFR.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N. A.
RC      STRAIN=C57BL/6J; TISSUE=LIVER;
RA      Reiter J.L., Theadgill D.W., Danielsen A.J., Schell C.,
RA      Lampland A.L., Balasubramaniam S., Crossley T.O., Magnuson T.R.,
RA      Maible N.J.;
RT      *Alternative transcripts from the Human and Mouse EGFR Genes Encode
RT      Cysteine-Terminal Truncated Receptors.
RL      Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
RN      [2]
RP      SEQUENCE FROM N. A.
RC      STRAIN=C3H/101, 129/SVJ, AND 129/SVEV/TAC;
RA      Reiter J.L., Theadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA      Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA      Balasubramaniam S., Crossley T.O., Magnuson T.R., James C.D.,
RA      Maible N.J.;
RT      *Comparative genomic sequence analysis and isolation of human and
RT      mouse alternative Egfr transcripts encoding truncated receptor
RT      isoforms.
RL      Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN      [3]
RP      SEQUENCE FROM N. A.
RC      STRAIN=C57BL/6J; TISSUE=LIVER;
RC      MEDLINE=21085660; PubMed=11217851;
RA      Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA      Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA      Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA      Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA      Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA      Fletschmann W., Gaasterland T., Gissi C., King B., Kochia H.,
RA      Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA      Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA      Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA      Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA      Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA      Gustincich S., Hill D., Hofmann M., Hume D.A., Kanliya M., Lee N.H.,
RA      Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombetars P.,
RA      Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA      Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA      Suzuki H., Toyooka K., Wang K.H., Weitz C., Wittaker C., Wilming L.,
RA      Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,
RA      Hayashizaki Y.;
RT      *Functional annotation of a full-length mouse cDNA collection.
RL      Nature 409:685-690(2001).
DR      EMBL; AF124513; AAD44149.1; -.
DR      EMBL; AF275366; AAG28047.1; -.
DR      EMBL; AF275364; AAG28047.1; JOINED.
DR      EMBL; AF275365; AAG28047.1; JOINED.
DR      EMBL; AK004944; BAB23688.1; -.
DR      EMBL; AK004883; BAB23641.1; -.
DR      EMBL; AK004911; BAB23662.1; -.
DR      MGI; 95294; Egfr.
DR      InterPro; IPR000494; EGFR_L.
DR      InterPro; IPR002174; Furin-like.
DR      Pfam; PF00757; Furin-like; 1.
DR      Pfam; PF01030; Recep_L_domain; 2.
DR      SMART; SM00261; FU; 3.
KW      Receptor.
SQ      SEQUENCE      655 AA; 72906 MW; 6B34063B1BC928CB CRC64;

Query Match      34.5%; Score 789; DB 11; Length 655;
Best Local Similarity 46.3%; Pred. No. 2e-60;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

```



```

QY 11 LLLALPPGAA--STQVCTGDMKRLRLPASPETHLMDLRHLHYOGGVVGNLELTPTN 68
DB 14 LITLCAAGGAALEEKVCOGTSNRLTQLGTFEDHFLSLQRMNNEVVLGNLEITYVRN 73
QY 69 ASLSFLDIOEVGVYLLAHNOVROVPLQRLRIYRGTOLEFEDNVALAVLDGDPPLNNTTP 128
DB 74 YDLSFLKTIQEVAGVYLLALMTVERIPLENQIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDI----FHKNNOLATLI 184
DB 125 -YGTNRKTLRELPMRNLOEILLGAVRFSNNPILCMNDTIQMRDIQVNFMSMSMDL--- 180
QY 185 DTNRSRACHPCSPMKSGKRCGSESEDCOSLRTVYACAGCA-RCKGPLPTDCCHEQCAAG 243
DB 181 -QSHSPSCPKCDPSCPNCSGCGEENCQKLTIIICAOQCSHRCGRSPDCHNOQCAAG 239
QY 244 CTGPKHSDCLALHFHNSGICELHCPALVTYNTDFFESMPNREGRTYGASCVTACPYNY 303
DB 240 CTGPRESCLVCOQRFQDEATKDCIPPLMLNPTTYQMDVNEPKGYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCAR 340
DB 300 VVTDHSGCVACGPDYEV-EEDGIRKCKKCDGCRK 335

RESULT 10
Q9EP98 PRELIMINARY; PRT; 1210 AA.
AC Q9EP98;
DT 01-MAR-2001 (Tremblrel, 16, Created)
DT 01-MAR-2001 (Tremblrel, 16, Last sequence update)
DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR ISOFORM 1.
GN EGFR.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OC NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C3H/101, 129/SVJ, 129/SVEVTA;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Sinclair C.S., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J;
RA Reiter J.L., Threadgill D.W., Eley G.D., Strunk K.E., Danielsen A.J.,
RA Schehl C., Pearsall R.S., Green P.J., Yee D., Lampland A.L.,
RA Balasubramanian S., Crossley T.O., Magnuson T.R., James C.D.,
RA Mahle N.J.;
RT "Comparative genomic sequence analysis and isolation of human and
RT mouse alternative Egr transcripts encoding truncated receptor
RT isoforms."
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF275366; AAG28045.1; JOINED.
DR EMBL: AF275365; AAG28045.1; JOINED.
DR EMBL: AF275367; AAG24386.1; -.
DR MGI: 95294; Egr.
DR InterPro: IPR000494; EGFR_L.
DR InterPro: IPR000719; Euk_PKinase.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002290; Ser_thr_kin_actsite.
DR InterPro: IPR001245; Tyr_kin.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF00069; pkinase; 1.

```

```

DR Pfam: PF01030; Recep_L_domain; 2.
DR PRINTS: PR00109; TYRKINASE.
DR SMART: SM00261; FU; 5.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00219; TYRK; 1.
DR PROSITE: PS00107; PROTEIN KINASE_ATP; 1.
DR PROSITE: PS00111; PROTEIN KINASE_DOM; 1.
DR ATP-binding; Receptor; Transferase.
DR SEQUENCE 1210 AA; 134841 MW; 62CD021C9DE32E18 CRC64;

Query Match 34.5%; Score 789; DB 11; Length 1210;
Best Local Similarity 46.3%; Pred. No. 4e-60;
Matches 156; Conservative 43; Mismatches 116; Indels 22; Gaps 6;

QY 11 LLLALPPGAA--STQVCTGDMKRLRLPASPETHLMDLRHLHYOGGVVGNLELTPTN 68
DB 14 LITLCAAGGAALEEKVCOGTSNRLTQLGTFEDHFLSLQRMNNEVVLGNLEITYVRN 73
QY 69 ASLSFLDIOEVGVYLLAHNOVROVPLQRLRIYRGTOLEFEDNVALAVLDGDPPLNNTTP 128
DB 74 YDLSFLKTIQEVAGVYLLALMTVERIPLENQIIRGNALYENTYALALISN----- 124
QY 129 VTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDI----FHKNNOLATLI 184
DB 125 -YGTNRKTLRELPMRNLOEILLGAVRFSNNPILCMNDTIQMRDIQVNFMSMSMDL--- 180
QY 185 DTNRSRACHPCSPMKSGKRCGSESEDCOSLRTVYACAGCA-RCKGPLPTDCCHEQCAAG 243
DB 181 -QSHSPSCPKCDPSCPNCSGCGEENCQKLTIIICAOQCSHRCGRSPDCHNOQCAAG 239
QY 244 CTGPKHSDCLALHFHNSGICELHCPALVTYNTDFFESMPNREGRTYGASCVTACPYNY 303
DB 240 CTGPRESCLVCOQRFQDEATKDCIPPLMLNPTTYQMDVNEPKGYSFGATCVKCKPRNY 299
QY 304 LSTDVGSCTLVCPPLHNOEVTAEADGTORCEKSKPCAR 340
DB 300 VVTDHSGCVACGPDYEV-EEDGIRKCKKCDGCRK 335

RESULT 11
Q9OX70 PRELIMINARY; PRT; 1209 AA.
AC Q9OX70;
DT 01-MAY-2000 (Tremblrel, 13, Created)
DT 01-MAY-2000 (Tremblrel, 13, Last sequence update)
DT 01-JUN-2001 (Tremblrel, 17, Last annotation update)
DE EPIDERMAL GROWTH FACTOR RECEPTOR.
GN EGFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OC NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A., Harris J., Raymond V.W., Blasband A.J., Lee D.C.,
RA Earp H.S.;
RT "A truncated, secreted form of the epidermal growth factor receptor is
RT encoded by an alternatively spliced transcript in normal rat tissue."
RL Mol. Cell. Biol. 10:2973-2982(1990).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Petch L.A.;
RL Submitted (NOV-1991) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=FISHER; TISSUE=LIVER;
RA Gutridge K., Dawson T.L., Earp H.S.;
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: M37394; AAF14008.1; -.

```


Best Local Similarity 45.7%; Pred. No. 3.4e-59;
Matches 154; Conservative 43; Mismatches 118; Indels 22; Gaps 6;

OY 11 LLLALLPPGAA--STOVCTGDMKRLRSPETHLMDLRLHYOGCOVQGNLELYLPN 68
DB 14 LLLALCAAGALEEKVKCGTSGTNRLLTQLTFEDHFLSLQRMNCEVYGNLEITYVORN 73
OY 69 ASLSFADIDIOEVQGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVDNGDPLNTPR 128
DB 74 YDLSTLKTIOEAGYFLNLNTERIPSEDLOIRGNALYEMTYALATLSN----- 124
OY 129 VTGASPGRLQLRLSLTEILKGVLIQNPOLCYODTILMKDI---FHKNNQALATLI 184
DB 125 -GTRNRTGLRELPMRNLDEILIGAVRFSNNPILCNMNDTIQMRDIQVNVFMSMSMDL--- 180
OY 185 DTRNRACHPCSPMCKSGRCMGESESDQSLRTYVAGGCA-RCKGRLPTDCCHEQCAAG 243
DB 181 -QSHFSSCPKCPSCPNCSGCGEENCQKTKRIICAQOCSHRGRSPSDCCHEQCAAG 239
OY 244 CTGPKHSDCLACIHFHNSGICELHCPALVYNTDFESMPNREGRTFGASCVTACPNY 303
DB 240 CTGPKHSDCLACIHFHNSGICELHCPALVYNTDFESMPNREGRTFGASCVTACPNY 299
OY 304 ISTDVSGCTLVCPRLHNOEVTADGTQRCCKSPCAR 340
DB 300 VYTDHSGCVACGPDYEV-EEDGIRKCKKCDGPKRK 335

RESULT 14
OYBUD7 PRELIMINARY; PRT: 331 AA.
ID OYBUD7;

AC 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE SIMILAR TO V-ERB-B2 AVIAN ERYTHROBLASTIC LEUKEMIA VIRAL ONCOGENE
DE HOMOLOG 3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=CHORIOCARINOMA;
RA Strausberg R.;
RL Submitted (Feb-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002706; AAH02706.1;--
SQ SEQUENCE 331 AA: 36489 MW: 45B8EBE683FE7E8 CRC64;

Query Match 32.1%; Score 734; DB 4; Length 331;
Best Local Similarity 44.0%; Pred. No. 5.6e-56;
Matches 147; Conservative 49; Mismatches 120; Indels 18; Gaps 8;

OY 10 GLLALLPPGAA--STOVCTGDMKRLRSPETHLMDLRLHYOGCOVQGNLELYLPN 67
DB 11 GLTFSLANGSEVNSQAVCPGLNLSTVGDENQYOTLYKIERCEVYGNLELYLICH 70
OY 68 NASLSFADIDIOEVQGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVDNGDPLNTP 127
DB 71 NADLSFADIDIOEVQGYVLLAHNOVROVPLQRLRIVRGTOLEEDNYALAVDNGDPLNTP 125
OY 128 PVTGASPGRLQLRLSLTEILKGVLIQNPOLCYODTILMKDI-FHKNNQALATLI 187
DB 126 ---NSSHALRQLRLTQLTEILSGVYLEKNDKLCMDPTIDMRDIDVRD--AAEIVKD 178
OY 188 RSRACHPCSPMCKSGRCMGESESDQSLRTYVAGGCA-ARCKGRLPTDCCHEQCAAGTG 246
DB 179 NGRSCRPCHVEYKRG-RCMGRSESDQSLRTYVAGGCA-ARCKGRLPTDCCHEQCAAGTG 237
OY 247 PHSKDLACIHFHNSGICELHCPALVYNTDFESMPNREGRTFGASCVTACPNY 306
DB 238 PDFTDCAFACRHHNDGAGCVPRLVYNNKLTFLQLEPNPHTKYQYGVAVASCPHNFV-V 296

OY 307 DVGSGCTVCPILHNOEVTADGTQRCCKSPCAR 340
DB 297 DOTSCYRACPPDKMEVD-KNGLKCEPCCGGLCPK 329

RESULT 15
OYB66 PRELIMINARY; PRT: 149 AA.
ID OYB66;
AC 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE RECEPTOR TYROSINE KINASE ERBB2 (FRAGMENT).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN (1)
RP SEQUENCE FROM N.A.
RA Klonisch T., Wolf P., Hombach-Klonisch S., Vogt S., Kuechenhoff A.,
RA Telens F., Fischer B.;
RT "Erdb genes and epidermal growth factor- (EGF-) like ligands in the
RT peri-implantation rabbit uterus and blastocyst."
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF333178; AAK14371.1;--
KM Kinase.
FT NON-TER 1 1
FT NON-TER 149 149
SQ SEQUENCE 149 AA: 16240 MW: 7CB3792A54FC49BA CRC64;

Query Match 31.6%; Score 723; DB 6; Length 149;
Best Local Similarity 81.2%; Pred. No. 2e-55;
Matches 121; Conservative 8; Mismatches 20; Indels 0; Gaps 0;

OY 158 NPQCYOPTILMKDI-FHKNNQALATLIIDNRSRACHPCSPMCKSGRCMGESESDQSLTR 217
DB 1 NPQCYOPTILMKDI-FHKNNQALATLIIDNRSRACHPCSPMCKSGRCMGESESDQSLTR 60
OY 218 TVGAGCARCKRPLPTDCCHEQCAAGTGPKHSDCLACIHFHNSGICELHCPALVYNTD 277
DB 61 TICAGGACARCKRPLPTDCCHEQCAAGTGPKHSDCLACIHFHNSGICELHCPALVYNTD 120
OY 278 TPESMPNREGRTFGASCVTACPNY 306
DB 121 TPESMPNREGRTFGASCVTACPNY 149

Search completed: April 11, 2002, 09:35:08
Job time: 325 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: April 11, 2002, 09:29:35 ; Search time 20.27 Seconds
(without alignments)
465.164 Million cell updates/sec

Title: US-09-234-208b-2

Sequence: 1 MELALCRWGLLLALLPPGA.....VGRCPDDAHVAVLSRYEG 419

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/PTVUS.COMB.pep:*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1878	82.1	782	US-09-146-283-4	Sequence 4, Appli
2	1878	82.1	782	US-08-579-823A-4	Sequence 4, Appli
3	1878	82.1	782	US-09-344-195-4	Sequence 4, Appli
4	1878	82.1	1255	US-08-467-083-68	Sequence 68, Appli
5	1878	82.1	1255	US-08-414-417B-68	Sequence 68, Appli
6	1878	82.1	1255	US-08-484-438-8	Sequence 8, Appli
7	1878	82.1	1255	US-08-486-348A-68	Sequence 68, Appli
8	1878	82.1	1255	US-08-625-101-2	Sequence 2, Appli
9	1878	82.1	1255	US-08-468-545B-68	Sequence 68, Appli
10	1878	82.1	1255	US-08-356-786-2	Sequence 2, Appli
11	1878	82.1	1255	US-08-466-680B-68	Sequence 68, Appli
12	1769	77.4	624	US-08-422-108-1	Sequence 1, Appli
13	793	34.7	644	US-08-336-708A-9	Sequence 9, Appli
14	793	34.7	1210	US-08-484-438-7	Sequence 7, Appli
15	793	34.7	1210	US-08-475-035-4	Sequence 4, Appli
16	775	33.9	911	US-08-484-438-10	Sequence 10, Appli
17	775	33.9	1058	US-08-484-438-4	Sequence 4, Appli
18	775	33.9	1308	US-08-484-438-2	Sequence 2, Appli
19	735.5	32.2	1342	US-07-978-895-4	Sequence 4, Appli
20	735.5	32.2	1342	US-08-484-438-9	Sequence 9, Appli
21	735.5	32.2	1342	US-08-473-119-4	Sequence 4, Appli
22	735.5	32.2	1342	US-08-475-352-4	Sequence 4, Appli
23	734	32.1	1343	5183884-4	Patent No. 5183884
24	493	21.6	97	US-08-421-356-3	Sequence 3, Appli
25	264.5	11.6	1382	US-08-737-715-2	Sequence 2, Appli
26	257.5	11.3	516	US-08-746-559A-4	Sequence 4, Appli
27	257.5	11.3	1367	US-08-249-687C-2	Sequence 2, Appli

28	257.5	11.3	1367	US-08-625-819-2	Sequence 2, Appli
29	257.5	11.3	1367	US-08-746-559A-2	Sequence 2, Appli
30	241.5	10.6	486	US-08-746-559A-5	Sequence 5, Appli
31	210.5	9.2	383	US-08-857-076-105	Sequence 105, Appli
32	203	8.9	11724	US-08-857-076-12	Sequence 12, Appli
33	196	8.6	366	US-08-857-076-103	Sequence 103, Appli
34	184.5	8.1	370	US-08-857-076-104	Sequence 104, Appli
35	147.5	6.4	381	US-08-857-076-106	Sequence 106, Appli
36	142	6.2	1940	US-08-644-271-30	Sequence 30, Appli
37	131.5	5.7	799	US-08-525-940-23	Sequence 23, Appli
38	131.5	5.7	799	US-08-976-838-23	Sequence 23, Appli
39	131.5	5.7	881	US-08-525-940-21	Sequence 21, Appli
40	131.5	5.7	881	US-08-976-838-21	Sequence 21, Appli
41	131.5	5.7	915	US-08-525-940-18	Sequence 18, Appli
42	131.5	5.7	915	US-08-976-838-18	Sequence 18, Appli
43	126	5.5	288	US-08-368-852-15	Sequence 15, Appli
44	124	5.4	288	US-08-525-940-15	Sequence 15, Appli
45	124	5.4	288	US-08-976-838-15	Sequence 15, Appli

ALIGNMENTS

RESULT 1
US-09-146-283-4
; Sequence 4, Application US/09146283
; Patent No. 5976546
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,283
; FILING DATE: 03-SEPT-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010.21
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
; US-09-146-283-4

Query Match 82.1%; Score 1878; DB 2; Length 782;
Best Local Similarity 83.0%; Pred. No. 5,1e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

OY 1 MELAAICRMGILLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
DB 1 MELAAICRMGILLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
OY 61 ELTYLPTNASLSFLDIOGVGVYLIANOVQVPLQRLRIYRGTOLEFEDNALAVLNDG 120
DB 61 ELTYLPTNASLSFLDIOGVGVYLIANOVQVPLQRLRIYRGTOLEFEDNALAVLNDG 120
OY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDTNRBRACHPCSPMKGKRCWSESEDCOSLTRVCAGGACARCKPLPTDCHEQC 240
DB 181 LTLIDTNRBRACHPCSPMKGKRCWSESEDCOSLTRVCAGGACARCKPLPTDCHEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
OY 301 YNVLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR-----GTHSLPRPAVPVP 355
DB 301 YNVLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR-----GTHSLPRPAVPVP 355
OY 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
OY 361 IOEFACCKRIFGSLAPLPSPFDDPASNT---APLOPQLOVFEITLGYIISAWPD 417
DB 361 IOEFACCKRIFGSLAPLPSPFDDPASNT---APLOPQLOVFEITLGYIISAWPD 417
OY 406 --PDAHVAVNLSTRYEG 419
DB 406 --PDAHVAVNLSTRYEG 419
OY 418 SLPDLSVFONLOVIRG 433
DB 418 SLPDLSVFONLOVIRG 433

```

RESULT 2

```

US-08-579-823A-4
; Sequence 4, Application US/08579823A
; Patent No. 6080409
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Composition and Method
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/579,823A
; FILING DATE: 03-DEC-1998
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Judge, Linda R.
; REGISTRATION NUMBER: 42,702
; REFERENCE/DOCKET NUMBER: 7636-0010
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 650-324-0880
; TELEFAX: 650-324-0960
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 782 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO

```

```

; ORIGINAL SOURCE:
; ORGANISM: homo sapiens
; INDIVIDUAL ISOLATE: GM-CSF-Her-2 fusion protein; Fig. 8
US-08-579-823A-4

```

```

Query Match      82.1%; Score 1878; DB 3; Length 782;
Best Local Similarity 83.0%; Pred. No. 5,1e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```

```

OY 1 MELAAICRMGILLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
DB 1 MELAAICRMGILLALLPFGAASVOTCTGDMKRLPASPETHLMDLRHLHYGCOVVOGNTL 60
OY 61 ELTYLPTNASLSFLDIOGVGVYLIANOVQVPLQRLRIYRGTOLEFEDNALAVLNDG 120
DB 61 ELTYLPTNASLSFLDIOGVGVYLIANOVQVPLQRLRIYRGTOLEFEDNALAVLNDG 120
OY 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTPTVTGASPGGLRELQRLSLTEILKGVLIQIRNPOLCYODTILMKDIFHKNNOLA 180
OY 181 LTLIDTNRBRACHPCSPMKGKRCWSESEDCOSLTRVCAGGACARCKPLPTDCHEQC 240
DB 181 LTLIDTNRBRACHPCSPMKGKRCWSESEDCOSLTRVCAGGACARCKPLPTDCHEQC 240
OY 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
DB 241 AAGCTGPKHSDCLACHFNHSGICELHCPALVTYNTDFESMPNPGRTTFCASCTACP 300
OY 301 YNVLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR-----GTHSLPRPAVPVP 355
DB 301 YNVLSTDVGSCTLVCPHLHNOEVAEDGTORCEKSKPCAR-----GTHSLPRPAVPVP 355
OY 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 356 LRMQPG--PAHPYLSFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
OY 361 IOEFACCKRIFGSLAPLPSPFDDPASNT---APLOPQLOVFEITLGYIISAWPD 417
DB 361 IOEFACCKRIFGSLAPLPSPFDDPASNT---APLOPQLOVFEITLGYIISAWPD 417
OY 406 --PDAHVAVNLSTRYEG 419
DB 406 --PDAHVAVNLSTRYEG 419
OY 418 SLPDLSVFONLOVIRG 433
DB 418 SLPDLSVFONLOVIRG 433

```

RESULT 3

```

US-09-344-195-4
; Sequence 4, Application US/09344195
; Patent No. 6210662
; GENERAL INFORMATION:
; APPLICANT: Laus, Reiner
; APPLICANT: Ruegg, Curtis L.
; APPLICANT: Wu, Hongyu
; TITLE OF INVENTION: Immunostimulatory Compositions
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Ave. Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/344,195
; FILING DATE: 24-Jun-1999
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/146,283
; FILING DATE: 03-SEPT-1998
; ATTORNEY/AGENT INFORMATION:

```



```

APPLICANT: Disis, Mary L.
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/414,417B
FILING DATE: 31-MAR-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-414-417B-68

Query Match      82.1%; Score 1878; DB 1; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALALCRWGLLLALLPPGAASSTOVCTGDMKRLRPASPEHLDMLRHLRYGCGVQVGNL 60
DB 1 MELALALCRWGLLLALLPPGAASSTOVCTGDMKRLRPASPEHLDMLRHLRYGCGVQVGNL 60
QY 61 ETTYLPNASTSLFDIODIEVOGYVLIANNOYROYVLRIRIVRGTOLEFEDNYALAVLDNG 120
DB 61 ETTYLPNASTSLFDIODIEVOGYVLIANNOYROYVLRIRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELDRLSLTEILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELDRLSLTEILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180
QY 181 LTLIDTNSRACHPCSPMKSGRCGSESEDCQSLTRTYVACGACRCKGPLETDCCHQC 240
DB 181 LTLIDTNSRACHPCSPMKSGRCGSESEDCQSLTRTYVACGACRCKGPLETDCCHQC 240
QY 241 AAGCTGPKHSDCLACLNHNSGICELHCPALVTYNTDFFESPNREGRTFEGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLNHNSGICELHCPALVTYNTDFFESPNREGRTFEGASCVTACP 300
QY 301 YNYLSTDVGSCITVCPPLHNOEYTAEDGTQRCCKSPCAR----GTHSLLRPPAAVPP 355
DB 301 YNYLSTDVGSCITVCPPLHNOEYTAEDGTQRCCKSPCAR----GTHSLLRPPAAVPP 355
QY 356 LRMQPG--PAIRVSLFLRPMNDLVAFSLPLAPLSPTSVPL-----SVYSGRGPD 405
DB 356 LRMQPG--PAIRVSLFLRPMNDLVAFSLPLAPLSPTSVPL-----SVYSGRGPD 405
QY 406 --PDAHVAVNLSTRYEG 419
DB 406 --PDAHVAVNLSTRYEG 419
QY 418 SLPLDLSVFQNLQVING 433
DB 418 SLPLDLSVFQNLQVING 433

```

```

RESULT 6
US-08-484-438-8
Sequence 8, Application US/08484438
Patent No. 581098 5780031
Patent No. 581098 5780031
GENERAL INFORMATION:
APPLICANT: Plowman, Gregory D.
APPLICANT: Culouscou, Jean-Michel
APPLICANT: Shoyab, Mohammed
APPLICANT: Stegall, Clay B.
APPLICANT: Hellstr m, Inggerd
APPLICANT: Hellstr m, Karl E.
TITLE OF INVENTION: HER4 HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennte & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistrock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-438-8

Query Match      82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALALCRWGLLLALLPPGAASSTOVCTGDMKRLRPASPEHLDMLRHLRYGCGVQVGNL 60
DB 1 MELALALCRWGLLLALLPPGAASSTOVCTGDMKRLRPASPEHLDMLRHLRYGCGVQVGNL 60
QY 61 ETTYLPNASTSLFDIODIEVOGYVLIANNOYROYVLRIRIVRGTOLEFEDNYALAVLDNG 120
DB 61 ETTYLPNASTSLFDIODIEVOGYVLIANNOYROYVLRIRIVRGTOLEFEDNYALAVLDNG 120
QY 121 DPLNNTTPTVTGASPGGLRELDRLSLTEILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVTGASPGGLRELDRLSLTEILKGGVLIQRNPQLCYODTILMKDIFHKNNOLA 180

```


QY 181 LTLIDNRSRACHPCSPMKGSRGSESESDCOSLTRVCAGGACRCKPPLPTDCHEOC 240
DB 181 LTLIDNRSRACHPCSPMKGSRGSESESDCOSLTRVCAGGACRCKPPLPTDCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
QY 301 YNVLSTDVSGCTLVCPPLHNOEYTAEDGTORCEKSPCAR-----GTHSLFRPAAVPPV 355
DB 301 YNVLSTDVSGCTLVCPPLHNOEYTAEDGTORCEKSPCAR-----GTHSLFRPAAVPPV 355
QY 356 LHMOPG--PAHVLSPFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 356 LHMOPG--PAHVLSPFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
QY 361 IOEFGCKKIFGSLAFPLPSFDGDPASNT---APLOPELOQVETLEITGTYISAMPD 417
DB 361 IOEFGCKKIFGSLAFPLPSFDGDPASNT---APLOPELOQVETLEITGTYISAMPD 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 SLPLDSVFQNLQVIRG 433

RESULT 7
US-08-486-348A-68
; Sequence 68, Application US/08486348A
; Patent No. 5846538
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
; TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
; TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 69
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Seed and Berry LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: US
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/486,348A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
US-08-486-348A-68

Query Match 82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9,5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 MELALCGMGILLALPPRAASTOVCTGDMKRLPASPETHLDMRLHLYGCGVQVQGNL 60
DB 1 MELALCGMGILLALPPRAASTOVCTGDMKRLPASPETHLDMRLHLYGCGVQVQGNL 60
QY 61 ELTYLPTNASISFLDIOIEVOGVYLIANHNOVRQVPLQRLIRIVRGTLQFEDNYALAVLDNG 120

DB 61 ELTYLPTNASISFLDIOIEVOGVYLIANHNOVRQVPLQRLIRIVRGTLQFEDNYALAVLDNG 120
QY 121 DELNNTPTPTGASPGGLRLOLRSLTEILKGVLIORNPOLCYODTILMKDIFHKNNOIA 180
DB 121 DELNNTPTPTGASPGGLRLOLRSLTEILKGVLIORNPOLCYODTILMKDIFHKNNOIA 180
QY 181 LTLIDNRSRACHPCSPMKGSRGSESESDCOSLTRVCAGGACRCKPPLPTDCHEOC 240
DB 181 LTLIDNRSRACHPCSPMKGSRGSESESDCOSLTRVCAGGACRCKPPLPTDCHEOC 240
QY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
DB 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTFGASCVTACP 300
QY 301 YNVLSTDVSGCTLVCPPLHNOEYTAEDGTORCEKSPCAR-----GTHSLFRPAAVPPV 355
DB 301 YNVLSTDVSGCTLVCPPLHNOEYTAEDGTORCEKSPCAR-----GTHSLFRPAAVPPV 355
QY 356 LHMOPG--PAHVLSPFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
DB 356 LHMOPG--PAHVLSPFLRSPMDLVSAFYSLPLAPLSPTSVPI-----SPVSGRGPD 405
QY 361 IOEFGCKKIFGSLAFPLPSFDGDPASNT---APLOPELOQVETLEITGTYISAMPD 417
DB 361 IOEFGCKKIFGSLAFPLPSFDGDPASNT---APLOPELOQVETLEITGTYISAMPD 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 SLPLDSVFQNLQVIRG 433

RESULT 8
US-08-625-101-2
; Sequence 2, Application US/08625101
; Patent No. 5869445
; GENERAL INFORMATION:
; APPLICANT: Cheever, Martin A.
; APPLICANT: Disis, Mary L.
; TITLE OF INVENTION: COMPOUNDS FOR ELICITING OR ENHANCING IMMUNE
; TITLE OF INVENTION: REACTIVITY TO HER-2/neu PROTEIN FOR PREVENTION
; TITLE OF INVENTION: OR TREATMENT OF MALIGNANCIES IN WHICH THE HER-2/neu
; TITLE OF INVENTION: ONCOGENE IS ASSOCIATED
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: Washington
; COUNTRY: USA
; ZIP: 98104-7092
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/625,101
; FILING DATE: 01-APR-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sharkey, Richard G.
; REGISTRATION NUMBER: 32,629
; REFERENCE/DOCKET NUMBER: 920010.448C7
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1255 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-625-101-2

Query Match 82.1%; Score 1878; DB 2; Length 1255;

Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```
OY 1 MELAALCRGGLLALLPPEAASVCTGTDMLRLPASETHLMDLRHLYGCGVYVGNL 60
Db 1 MELAALCRGGLLALLPPEAASVCTGTDMLRLPASETHLMDLRHLYGCGVYVGNL 60
OY 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFEDNALAVLDNG 120
Db 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFEDNALAVLDNG 120
OY 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNOLA 180
Db 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNOLA 180
OY 181 LTLIDTNRBRACHPCSPMKGSRGCGESSEDCOSLTRVCAGGACRCKPLPTDCCHQC 240
Db 181 LTLIDTNRBRACHPCSPMKGSRGCGESSEDCOSLTRVCAGGACRCKPLPTDCCHQC 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCTYAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCTYAC 300
OY 301 YVLTSDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR---GTHSLRPPAAVVP 355
Db 301 YVLTSDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR---GTHSLRPPAAVVP 355
OY 356 LKMQRG--PAHVLSTLRSMPLVSFSLPLAPLSPTVPI-----SPVSVGRGD 405
Db 356 LKMQRG--PAHVLSTLRSMPLVSFSLPLAPLSPTVPI-----SPVSVGRGD 405
OY 406 --PDAHVAVNLRSYEG 419
Db 406 --PDAHVAVNLRSYEG 419
OY 418 SLPDLSVFONLOYIRG 433
Db 418 SLPDLSVFONLOYIRG 433
```

RESULT 9

US-08-468-545B-68

Sequence 68, Application US/08468545B

Patent No. 5876712

GENERAL INFORMATION:

APPLICANT: Cheever, Martin A.

APPLICANT: Disis, Mary L.

TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN

TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE

TITLE OF INVENTION: HER-2/neu ONCOGENE IS ASSOCIATED

NUMBER OF SEQUENCES: 69

CORRESPONDENCE ADDRESS:

ADDRESSEE: Seed and Berry LLP

STREET: 6300 Columbia Center, 701 Fifth Avenue

CITY: Seattle

STATE: Washington

COUNTRY: US

ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,545B

FILING DATE: 06-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Sharkey, Richard G.

REGISTRATION NUMBER: 32,629

REFERENCE/DOCKET NUMBER: 920010.448C5

TELECOMMUNICATION INFORMATION:

TELEPHONE: (206) 622-4900

TELEFAX: (206) 682-6031

INFORMATION FOR SEQ ID NO: 68:

SEQUENCE CHARACTERISTICS:

LENGTH: 1255 amino acids

TYPE: amino acid
TOPOLOGY: linear
US-08-468-545B-68

Query Match 82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

```
OY 1 MELAALCRGGLLALLPPEAASVCTGTDMLRLPASETHLMDLRHLYGCGVYVGNL 60
Db 1 MELAALCRGGLLALLPPEAASVCTGTDMLRLPASETHLMDLRHLYGCGVYVGNL 60
OY 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFEDNALAVLDNG 120
Db 61 ELTYLPTNLSFLDIOGVGVYLIANQVQVPLQRLRIYRGTOLEFEDNALAVLDNG 120
OY 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNOLA 180
Db 121 DELNNTPTVTGASPGGLRELQRLSLTEILKGGVLIQRNPOLCYQDTILMKDIFHKNOLA 180
OY 181 LTLIDTNRBRACHPCSPMKGSRGCGESSEDCOSLTRVCAGGACRCKPLPTDCCHQC 240
Db 181 LTLIDTNRBRACHPCSPMKGSRGCGESSEDCOSLTRVCAGGACRCKPLPTDCCHQC 240
OY 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCTYAC 300
Db 241 AAGCTGPKHSDCLACLFHNSGICELHCPALVTYNTDFESMPNPEGRTYFGASCTYAC 300
OY 301 YVLTSDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR---GTHSLRPPAAVVP 355
Db 301 YVLTSDVGSCTLVCPPLHNOEYTAEDGTORCEKSKPCAR---GTHSLRPPAAVVP 355
OY 356 LKMQRG--PAHVLSTLRSMPLVSFSLPLAPLSPTVPI-----SPVSVGRGD 405
Db 356 LKMQRG--PAHVLSTLRSMPLVSFSLPLAPLSPTVPI-----SPVSVGRGD 405
OY 406 --PDAHVAVNLRSYEG 419
Db 406 --PDAHVAVNLRSYEG 419
OY 418 SLPDLSVFONLOYIRG 433
Db 418 SLPDLSVFONLOYIRG 433
```

RESULT 10

US-08-356-786-2

Sequence 2, Application US/08356786

Patent No. 5877305

GENERAL INFORMATION:

APPLICANT: Huston, James S.

APPLICANT: Oppermann, Hermann

APPLICANT: Houston, L. L.

APPLICANT: Ring, David B.

TITLE OF INVENTION: Biosynthetic Binding Protein for Cancer

NUMBER OF SEQUENCES: 16

CORRESPONDENCE ADDRESS:

ADDRESSEE: Edmund R. Pletcher, Testa, Hurwitz, & Thibault

STREET: Exchange Place, 53 State Street

CITY: Boston

STATE: Massachusetts

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/356,786

FILING DATE:

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/831,967

FILING DATE: 06-FEB-1992

ATTORNEY/AGENT INFORMATION:
NAME: Pitcher, Edmund R.
REGISTRATION NUMBER: 27,829
REFERENCE/DOCKET NUMBER: CRP-053
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 248-7000
TELEFAX: (617) 248-7100
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-356-786-2

Query Match 82.1%; Score 1878; DB 2; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALCRWGLLALLPPGASTQVCTGDMKRLPASPEHLDMLRHLYGCCQVVGNTL 60
DB 1 METALCRWGLLALLPPGASTQVCTGDMKRLPASPEHLDMLRHLYGCCQVVGNTL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYLIAHNOVQVPLQRLRIYRGQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLFLQDIOEVQGYLIAHNOVQVPLQRLRIYRGQLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
QY 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFEESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFEESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
QY 366 LRMQPG--PAHPVLSFLRPMNDVSAFYSPLAPLSPTSVPL-----SPVSYGRGPD 405
DB 361 IQEFAGCKRKIFGSLAFLEPSFDGDPASNT---APLPQPOLQVFETLEITGYLYISAMPD 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 SLPLDLSVFQNLQVIRG 433

RESULT 11
US-08-466-680B-68
Sequence 68, Application US/08466680B
Patent No. 6075122
GENERAL INFORMATION:
APPLICANT: Cheever, Martin A.
APPLICANT: Disis, Mary L.
TITLE OF INVENTION: IMMUNE REACTIVITY TO HER-2/neu PROTEIN
TITLE OF INVENTION: FOR DIAGNOSIS AND TREATMENT OF MALIGNANCIES IN WHICH THE
NUMBER OF SEQUENCES: 69
CORRESPONDENCE ADDRESS:
ADDRESSEE: Seed and Berry LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: Washington
COUNTRY: US
ZIP: 98104-7092
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/466,680B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Sharkey, Richard G.
REGISTRATION NUMBER: 32,629
REFERENCE/DOCKET NUMBER: 920010.448C4
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 1255 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-466-680B-68

Query Match 82.1%; Score 1878; DB 3; Length 1255;
Best Local Similarity 83.0%; Pred. No. 9.5e-156;
Matches 362; Conservative 9; Mismatches 45; Indels 20; Gaps 5;

QY 1 METALCRWGLLALLPPGASTQVCTGDMKRLPASPEHLDMLRHLYGCCQVVGNTL 60
DB 1 METALCRWGLLALLPPGASTQVCTGDMKRLPASPEHLDMLRHLYGCCQVVGNTL 60
QY 61 ELTYLPTNASLFLQDIOEVQGYLIAHNOVQVPLQRLRIYRGQLFEDNYALAVLNG 120
DB 61 ELTYLPTNASLFLQDIOEVQGYLIAHNOVQVPLQRLRIYRGQLFEDNYALAVLNG 120
QY 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
DB 121 DPLNNTTPTVGTASPGGLRELDRLSLTEILKGVLIIQRNPOLCYQDTILMKDIFHKNNOLA 180
QY 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
DB 181 LTLIDTNSRACHPCSPCKSGRCWGESSEDCQSLTRTVAGGACRGCPLEPTDCCHQC 240
QY 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFEESMPNEGRYTFGASCVTACP 300
DB 241 AAGCTGPRHSDCLACIHNHSGICELHCPALVTYNTDFEESMPNEGRYTFGASCVTACP 300
QY 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
DB 301 YNYLSTDVGSCTLVCPPLHNOEYTAEDGTQRCCKSPCARVCYGLGMEHLREVRVAVTSAN 360
QY 366 LRMQPG--PAHPVLSFLRPMNDVSAFYSPLAPLSPTSVPL-----SPVSYGRGPD 405
DB 361 IQEFAGCKRKIFGSLAFLEPSFDGDPASNT---APLPQPOLQVFETLEITGYLYISAMPD 417
QY 406 --PDAHVAVNLSRYEG 419
DB 418 SLPLDLSVFQNLQVIRG 433

RESULT 12
US-08-422-108-1
Sequence 1, Application US/08422108
Patent No. 6015567
GENERAL INFORMATION:
APPLICANT: Huddiak, Robert M.
APPLICANT: Shepard, H. Michael
APPLICANT: Ullrich, Axel
TITLE OF INVENTION: HER2 EXTRACELLULAR DOMAIN
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco

STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WinPatIn (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/422,108
FILING DATE: 14-Apr-1995
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/355460
FILING DATE: 13-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/048346
FILING DATE: 15-APR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/354319
FILING DATE: 19-MAY-1989
ATTORNEY/AGENT INFORMATION:
NAME: Lee, Wendy M
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 554C2D2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 624 amino acids
TYPE: Amino Acid
TOPOLOGY: Linear
US-08-422-108-1

Query Match 77.4%; Score 1769; DB 3; Length 624;
Best Local Similarity 82.2%; Pred. No. 1,3e-146;
Matches 341; Conservative 9; Mismatches 45; Indels 20; Gaps 5;
QY 22 STQVCTGDTMKLRLPASPEHLMRLHLYQGGVQVGNLELYLPTNASLFLQDIQEVQ 81
DB 1 STQVCTGDTMKLRLPASPEHLMRLHLYQGGVQVGNLELYLPTNASLFLQDIQEVQ 60
QY 82 GYVLAHNOVROVPLQRLRIYVGTQLFEDNYALAVLDNDPLNNTPTVVGASPGGLREIQ 141
DB 61 GYVLAHNOVROVPLQRLRIYVGTQLFEDNYALAVLDNDPLNNTPTVVGASPGGLREIQ 120
QY 142 LRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNQLALTLDITNRSRACHPCSPMCKG 201
DB 121 LRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNQLALTLDITNRSRACHPCSPMCKG 180
QY 202 SRQWSESSDCCSLTRTYVAGGACARCKGPLPTDCCHQCCAGCTGPKHSDCLACHFNHS 261
DB 181 SRQWSESSDCCSLTRTYVAGGACARCKGPLPTDCCHQCCAGCTGPKHSDCLACHFNHS 240
QY 262 GICEHLCRALVYNTDTPESMPNPEGRTYFGASCYACRYNTLSDVSGCTLVCPHLANE 321
DB 241 GICEHLCRALVYNTDTPESMPNPEGRTYFGASCYACRYNTLSDVSGCTLVCPHLANE 300
QY 322 VTAEDGTORCECKSPCAR-----GNHSLRPRAVVPVLRMQPG--PAHPIVLSFLRPSW 374
DB 301 VTAEDGTORCECKSPCARVCTGLGMEHLREYRAVTSANIQEPAGCKITFGSLAFRPESE 360
QY 375 DLVSFAFSLPLAPLSPTSVPI-----SPYVSGRGPD--PDAAVAVMSRYEG 419
DB 361 DGDPRASMT---APLQPEQVQVETLEETLEYLYISAMPDLSLDFVFOQLQYIRG 412

RESULT 13
US-08-336-708a-9
Sequence 9, Application US/08336708A

Patent No. 5521295
GENERAL INFORMATION:
APPLICANT: Pacificl, Robert E.
APPLICANT: Thomason, Arlen R.
APPLICANT: Chang, Ming-Shi
TITLE OF INVENTION: Hybrid Receptor Molecules
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Amgen Inc.
STREET: 1640 Dehavilland Drive
CITY: Thousand Oaks
STATE: California
COUNTRY: USA
ZIP: 91320-1789
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/336,708A
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Oleski, Nancy
REFERENCE/DOCKET NUMBER: A-241A
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 644 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-336-708a-9

Query Match 34.7%; Score 793; DB 1; Length 644;
Best Local Similarity 45.3%; Pred. No. 2.9e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;
QY 11 LIALLPQGA--STQVCTGDTMKLRLPASPEHLMRLHLYQGGVQVGNLELYLPTN 68
DB 14 LIAALCPASRALEEKVKVCGSTSNKLTQGTGFEDHFLSQRMNCEVYVGNLEITYVORN 73
QY 69 ASLSFLQDIQEVQGYVLAHNOVROVPLQRLRIYVGTQLFEDNYALAVLDNDPLNNTPT 128
DB 74 YDLSFLKTIQEVAGYVLAALNVERIPLENLIQIRGNMYENSVALAVLSND----- 126
QY 129 VTGASPGGLREQLRSLEILKGVLIQIRNPOLCYODTILMKDIFHKNNQLALTLDITNR 188
DB 127 ---AKTIGLKEPLMNLQELHGAVRFSNNPALCNVESIQWRDIYSSDFLSMSMDPQNH 183
QY 189 SRACHPCSPMCKGSRQWSESSDCCSLTRTYVAGGCA--RCKGPLPTDCCHQCCAGCTGP 247
DB 184 LSCCKCPSPCNGSCWGAEGENCOKLTKIICAQCCSGCRCKSPSDCHNCCAGCTGP 243
QY 248 KHSDDLACHFNHSGICEHLCRALVYNTDTPESMPNPEGRTYFGASCYACRYNTLSDV 307
DB 244 RESDCLVCKEFDEATCKDTCPLMLYNPTTYQMDVNPBGKYSFGATCVKCPRYVYTD 303
QY 308 VGSCTLVCPPLHNOEYTAEDGTORCECKSPCAR 340
DB 304 HGSYVRACGADSYEM--BEDGVAKCKCKGCPCKR 335

RESULT 14
US-08-484-438-7
Sequence 7, Application US/08484438
Patent No. 581098
Patent No. 581098 5780031
GENERAL INFORMATION:
APPLICANT: Plouman, Gregory D.
APPLICANT: Culouscou, Jean-Michel

APPLICANT: Shoyab, Mohammed
APPLICANT: Siegal, Clay B.
APPLICANT: Helistr m, Ingerd
APPLICANT: Helistr m, Karl E.
TITLE OF INVENTION: HERA HUMAN RECEPTOR TYROSINE KINASE
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/484,438
FILING DATE: 07-JUN-1995
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: 08/323,442
FILING DATE: 14-OCT-1994
APPLICATION NUMBER: US 08/150,704
FILING DATE: 10-NOV-1993
CLASSIFICATION: 530
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/981,165
FILING DATE: 24-NOV-1992
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Mistock, S. Leslie
REGISTRATION NUMBER: 18,872
REFERENCE/DOCKET NUMBER: 5624-230
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-8864/9741
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
STRANDEDNESS: unknown
TOPOLOGY: unknown
MOLECULE TYPE: protein
US-08-484-438-7

Query Match 34.7%; Score 793; DB 2; Length 1210;
Best Local Similarity 45.3%; Pred. No. 6,7e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

11 LLLALLPAGAA--STOVCTGDMKRLRLPASPTHLMLRLHYOGCOVVOGNLELTFTPTN 68
14 LLAALCPASRALEKKKVGCGTSNKLTOLGTFEDHFLSLQRMNNGCEVVLGNLEITTYOGRN 73
69 ASLSFLQDIOEVQGYVLAHNOYROYPLORLRIVGTQLFEDNYALAVLDNGDPLNNTTP 128
74 YDLSFLKTIOEVAGYVLAHNOYROYPLORLRIVGTQLFEDNYALAVLDNGDPLNNTTP 126
129 VTGASPGGLREIOLRSITELIKGVLIOGNPOLCYODITLMKDIPIKNNQALATLIDTNR 188
127 ---ANKTGLKEIPKRNLOETILHGAVERFSNNPALCNVESTIOMRDIVSSDPLSNMDFQNH 183
189 SRACHPCSPMKSGSRMGSSSEDQSLTFTVCAGGCA-RCKGPLPTDCCHEQCAACTGP 247
184 LSGCQKCDSPCGNSGCGWAGEENCKLTKIICAGQCSGRCRGSPPDCCHNGCAACTGP 243
248 KHSDDLACHFNHSGICELHCPALVTYNTDTEFSMPNPGRTYFGASCYATACPYNTLSTD 307
244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYOMDVNPGKYSFGATCVKCKPRNRYVVD 303

308 VGSCTLVCPPLHNOEVTAEQDTCRCEKSKPCAR 340
304 HGSVCVRACGADSYEM-BEDGVKCKRCEGPCRK 335

RESULT 15
US-08-475-035-4
Sequence 4, Application US/08475035
Patent No. 598553
GENERAL INFORMATION:
APPLICANT: KING, C. R.
APPLICANT: KRAUS, MATTHIAS H.
APPLICANT: AARONSON, STUART A.
TITLE OF INVENTION: HUMAN GENE RELATED TO BUT DISTINCT FROM
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: NEEDLE & ROSENBERG, P.C.
STREET: Suite 1200, 127 Peachtree Street
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/475,035
FILING DATE: 7 Jun 1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Pertyman, David G.
REGISTRATION NUMBER: 33,438
REFERENCE/DOCKET NUMBER: 1414.656
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1210 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-475-035-4

Query Match 34.7%; Score 793; DB 2; Length 1210;
Best Local Similarity 45.3%; Pred. No. 6,7e-61;
Matches 151; Conservative 50; Mismatches 118; Indels 14; Gaps 4;

11 LLLALLPAGAA--STOVCTGDMKRLRLPASPTHLMLRLHYOGCOVVOGNLELTFTPTN 68
14 LLAALCPASRALEKKKVGCGTSNKLTOLGTFEDHFLSLQRMNNGCEVVLGNLEITTYOGRN 73
69 ASLSFLQDIOEVQGYVLAHNOYROYPLORLRIVGTQLFEDNYALAVLDNGDPLNNTTP 128
74 YDLSFLKTIOEVAGYVLAHNOYROYPLORLRIVGTQLFEDNYALAVLDNGDPLNNTTP 126
129 VTGASPGGLREIOLRSITELIKGVLIOGNPOLCYODITLMKDIPIKNNQALATLIDTNR 188
127 ---ANKTGLKEIPKRNLOETILHGAVERFSNNPALCNVESTIOMRDIVSSDPLSNMDFQNH 183
189 SRACHPCSPMKSGSRMGSSSEDQSLTFTVCAGGCA-RCKGPLPTDCCHEQCAACTGP 247
184 LSGCQKCDSPCGNSGCGWAGEENCKLTKIICAGQCSGRCRGSPPDCCHNGCAACTGP 243
248 KHSDDLACHFNHSGICELHCPALVTYNTDTEFSMPNPGRTYFGASCYATACPYNTLSTD 307
244 RESDCLVCRKFRDEATCKDTCPLMLYNPTTYOMDVNPGKYSFGATCVKCKPRNRYVVD 303

OY 308 VGSCTLVCPHNEVTAEDGTORCEKCSKPCAR 340
||| | : | : || : | :
DB 304 HGSCTVRACGADSYEM-EDGVRKCKKCGPCRK 335

Search completed: April 11, 2002, 09:29:38
Job time: 75 sec